

## Prediction of River Discharge and Assessment its Relationship at Consecutive Hydrometric Stations Using GPR-EEMD Combined Techniques (Case Study: Housatonic River)

KIYOUMARSE ROUSHANGAR<sup>1\*</sup>, MASOUMEH CHAMANI<sup>1</sup>

1. Department of Water Engineering, Faculty of civil, Tabria University, Tabriz, Iran.

(Received: May. 15, 2019- Revised: July. 20, 2019- Accepted: July. 27, 2019)

### ABSTRACT

Accurate forecasting of river flow is one of the most important factors in surface water resources management, especially during flood and drought periods. In this research, the wavelet function and the ensemble empirical mode decomposition (EEMD), which are considered as soft computing tools, were used to derive the time series features, and the efficiency of the wavelet- Gaussian and the ensemble empirical mode decomposition-Gaussian models for predicting the discharge between the three consecutive stations located in the Housatonic river have been investigated. For this purpose, in the first step, the discharge of downstream stations is predicted by upstream stations using the Gaussian process regression model. Then, the discharge-stage time series was broken up by wavelet transform and ensemble empirical mode decomposition into cages, and these subclasses were introduced into the Gaussian process regression modeling to simulate the discharge-stage relationship. Also, the effect of each of the sub-series of ensemble empirical mode decomposition model (Residual and IMFs) was studied to improve predictive outcomes. It was observed that the most inefficient subseries in the ensemble empirical mode decomposition model is the residual subseries. The results indicate that wavelet compound techniques (DWT-GPR) and ensemble empirical mode decomposition (EEMD-GPR) have improved the results to a certain extent. As an example, for the test stage, the best prediction model of the second station, the combined model of ensemble empirical mode decomposition-Gaussian upgraded determination coefficient (DC) from 0.74 to 0.80 and the combined model of wavelet-Gaussian upgraded DC from 0.74 to 0.83.

**Keywords:** Discharge, Ensemble empirical mode decomposition, Discrete Wavelet Transform, Gaussian process regression, Sequential stations

## پیش‌بینی و ارزیابی ارتباط دبی رودخانه در ایستگاه‌های هیدرومتریک متوالی با استفاده از روش‌های ترکیبی GPR-EEMD (مطالعه موردی: رودخانه هوستونیک)

کیومرث روشنگر<sup>۱\*</sup>، معصومه چمنی<sup>۲</sup>

۱. گروه آب، دانشکده مهندسی عمران، دانشگاه تبریز، تبریز، ایران

(تاریخ دریافت: ۱۳۹۸/۲/۲۵ - تاریخ بازنگری: ۱۳۹۸/۴/۲۹ - تاریخ تصویب: ۱۳۹۸/۵/۵)

### چکیده

پیش‌بینی دقیق دبی در رودخانه‌ها، از مهم‌ترین مؤلفه‌های فرآیندهای هیدرولوژیکی و هیدرولیکی در مدیریت منابع آب، به ویژه در اتخاذ تدابیر مناسب در مواقع خشکسالی و بروز سیلاب است. در این تحقیق از تابع موجک و تجزیه مد تجربی یکپارچه که از ابزارهای محاسبات نرم محسوب می‌شوند، جهت استخراج ویژگی‌های سری زمانی استفاده گردیده و کارایی مدل‌های موجک-گوسین (GPR-DWT) و تجزیه مد تجربی یکپارچه-گوسین (GPR-EEMD) برای پیش‌بینی دبی بین سه ایستگاه متوالی رودخانه هوستونیک، واقع در آمریکا مورد بررسی قرار گرفته است. برای این منظور در گام اول، مقدار دبی ایستگاه پایین‌دست، توسط ایستگاه‌های بالادست با استفاده از مدل رگرسیون فرایند گاوسی پیش‌بینی شده است. سپس سری‌های زمانی دبی و اشل توسط تبدیل موجک و تجزیه مد تجربی یکپارچه به زیرسری‌هایی تجزیه گشته و این زیرسری‌ها جهت شبیه‌سازی رابطه دبی-اشل وارد مدل رگرسیون فرایند گاوسی شدند. همچنین تأثیر هر یک از زیرسری‌های روش تجزیه مد تجربی یکپارچه (Residual and IMFs) در نتایج پیش‌بینی، بررسی گردید. مشاهده گردید که ناکارآمدترین زیرسری در تجزیه مد تجربی یکپارچه، زیرسری باقیمانده (Residual) می‌باشد. نتایج حاکی از آن است که روش‌های ترکیبی موجک (GPR-DWT) و تجزیه مد تجربی یکپارچه (GPR-EEMD) تا حدود زیادی باعث بهبود نتایج گردیدند. به عنوان نمونه، برای مرحله آزمون مدل برتر پیش‌بینی دبی ایستگاه دوم، مدل تلفیقی تجزیه مد تجربی یکپارچه-گوسین  $DC=0/74$  را به  $DC=0/80$  و مدل تلفیقی موجک-گوسین  $DC=0/74$  را به  $DC=0/83$  ارتقاء داد.

**واژه‌های کلیدی:** دبی رودخانه، تجزیه مد تجربی یکپارچه، تبدیل موجک گسسته، رگرسیون فرایند گاوسی، ایستگاه‌های متوالی

### مقدمه

برای کاربردهای مختلف هیدرولوژیکی، مهندسی رودخانه و کنترل پروژه‌های منابع آبی، اطلاعات دقیق درباره میزان جریان در رودخانه‌ها بسیار مهم است. تا کنون روابط گوناگون و پیچیده-ای به منظور پیش‌بینی دبی رودخانه‌ها ارائه شده است. رابطه دبی-اشل در رودخانه یک پدیده بسیار پیچیده در علم هیدرولوژی است که شناخت آن در مدیریت و برنامه‌ریزی منابع آب از اهمیت بسیار زیادی برخوردار است. مدل‌سازی این فرآیند می‌تواند ضریب اطمینان بالایی در طراحی سازه‌های هیدرولیکی، ساماندهی رودخانه‌ها و برنامه‌ریزی برای سیستم‌های هشدار سیل را ممکن نماید.

در سال‌های اخیر روش‌های هوش مصنوعی (AI)<sup>۱</sup> توانایی

قابل ملاحظه‌ای در مدل‌سازی و پیش‌بینی سری‌های زمانی هیدرولوژیکی از خود نشان داده‌اند. این روش‌ها چون مبتنی بر داده هستند، نیاز به دانش اولیه نسبت به پدیده‌ی مورد مطالعه ندارند. مدل‌های مبتنی بر AI، ابزاری کارآمد برای کارکردن روی تعداد وسیعی از داده‌های دینامیکی و غیرخطی می‌باشند. به عنوان نمونه، Roushangar et al. (2013)، به منظور تخمین دبی روزانه رودخانه ونبار در استان آذربایجان شرقی از برنامه‌ریزی ژنتیک و شبکه عصبی مصنوعی استفاده کردند، و برای هر یک از روش‌ها بهترین ترکیب ورودی را تعیین کردند. همچنین نتایج به دست آمده، برتری برنامه‌ریزی ژنتیک را نسبت به شبکه عصبی مصنوعی نشان داد. Nouri et al. (2001)، با استفاده از مدل‌های SVM و ANN به پیش‌بینی دبی ماهانه پرداختند و

\* نویسنده مسئول: kroshangar@yahoo.com

1-Artificial Intelligence  
2-Support Vector Machine  
3-Artificial Neural Network

اساس مدل مرجع و مجموعه داده‌های ورودی و خروجی، روش صریح LGP با عملگرهای محاسباتی هم نتایج بهتری نسبت به مدل WANN از خود نشان داده است (Danandeh Mehr 2013). *et al.*، تیواری و آداموسکی دو نوع از مدل ترکیبی موجک-شبکه عصبی مصنوعی (WANN) برای پیش‌بینی تقاضای ماهانه و هفتگی آب را با داده‌های محدود در دسترس توسعه دادند. نتایج نشان می‌دهد که روش‌های ارائه‌شده قادرند تا دقت و اعتبار پیش‌بینی تقاضای آب را به وسیله ترکیب توانایی تبدیل موجک (WT) بهبود بخشند (Tiwari and Adamowski, 2014). آدارش و همکاران ابتدا با استفاده از آزمون ناپارامتریک من-کندال، روند سری‌های زمانی تغییرات سالیانه رسوب از خروجی ۷ ایستگاه اصلی حوضه در هند را تحلیل کردند. سپس روندهای غیر خطی در این سری‌ها را با استفاده از تجزیه مود تجربی (EMD) استخراج کردند. نتایج مطالعه تحلیل روند، کاهش تغییرات رسوب در ۵ حوضه از ۷ حوضه در هند را علی‌رغم افزایش میزان بارش‌ها در ۳ حوضه از آن‌ها نشان داد (Adarsh *et al.*, 2016). همچنین در مطالعه‌ای نشان دادند مدل ترکیبی موجک-شبکه عصبی مصنوعی در سری‌های زمانی هیدرولوژیکی دارای دقت بالاتری نسبت به مدل ساده شبکه عصبی مصنوعی است (Alizadeh *et al.*, 2017).

در تحقیق کنونی کارایی روش‌های رگرسیون فرایند گاوسی و تجزیه مد تجربی یکپارچه و تبدیل موجک به عنوان روش‌های نوین داده کاوی در تخمین دبی بین ایستگاهی مورد بررسی و ارزیابی قرار گرفت. ارزیابی مدل‌های ترکیبی با مقایسه و انتخاب مدل برتر بر اساس بهترین عملکرد کارایی انجام شد. همچنین تأثیر هر یک از زیرسری‌های تجزیه مد تجربی یکپارچه در نتایج پیش‌بینی بررسی گردید.

## مواد و روش

### منطقه مورد مطالعه

در این تحقیق از داده‌های روزانه دبی و اشل سال ۲۰۰۲ تا ۲۰۰۶ مربوط به ایستگاه‌های متوالی واقع در رودخانه هوستونیک که از سایت USGS استخراج شده است، استفاده گردید. در جدول (۱) اطلاعات مربوط به ایستگاه‌های هیدرومتریک متوالی رودخانه ارائه شده است. در شکل (۱) موقعیت منطقه مورد مطالعه آورده شده است. قابل ذکر است که جهت جریان رودخانه از ایستگاه اول به ایستگاه سوم بوده و فاصله ایستگاه‌ها از هم حدود ۵۰ کیلومتر می‌باشد.

نتیجه گرفتند که مدل SVM از دقت بالایی برخوردار است. کیشی و کوبنر با بررسی داده‌های سه ایستگاه رودخانه غزل ایرماق ترکیه و به کارگیری شبکه عصبی چند لایه پرسپترون، الگوریتم نیوتن و روش RBF نشان دادند که روش شبکه عصبی بهتر از روش رگرسیون بوده و می‌تواند به عنوان روشی مناسب برای پیش‌بینی دبی-اشل به کار برده شود (Kisi and 2009). Cobaner, بهزادی و همکاران در مطالعه‌ای، توسعه عملکرد روش‌های ماشین بردار پشتیبان و شبکه عصبی مصنوعی را برای پیش‌بینی یک روز جلوتر رواناب مورد بررسی قرار دادند. نتایج نشان داد که روش ماشین بردار پشتیبان از روش شبکه عصبی مصنوعی، پیش‌بینی سریع‌تر و دقیق‌تر از خود نشان داده است (Behzadi *et al.*, 2009). در تحقیقی از مدل SVR به منظور پیش‌بینی جریان رودخانه در مناطق مرطوب گرمسیری (مالزی) استفاده کردند و با تولید شش مدل مرتبط با SVR با تغییرات سطح آب و ارتفاع بارش در بالادست به پیش‌بینی جریان ساعتی رودخانه در مقطعی از سال ۲۰۱۱ در حوضه سلانگور ۶ پرداختند. نتایج آنها حاکی از کارایی SVR در پیش‌بینی جریان رودخانه بود (Sayam *et al.*, 2017). همچنین تحقیقات گسترده‌ای در زمینه مقایسه SVM با مدل‌های داده محور در هیدرولوژی انجام شده است. به طوری که برتری این مدل را نسبت به سایر مدل‌های داده‌محور خاص از قبیل MLR و برخی از ANN نشان می‌دهد. از همین رو در این میان بیشترین مقایسه مابین مدل ذکرشده و مدل ANN صورت گرفته است (Adnan *et al.*, 2017; Modaresi *et al.*, 2017).

با توجه به اینکه نتایج مدل‌های داده‌محور علاوه بر کمیت داده‌ها به کیفیت داده‌ها نیز بستگی دارد، در سال‌های اخیر محققان از روش‌های پیش‌پردازش نظیر تبدیل موجک برای بهبود نتایج شبیه‌سازی استفاده کردند. برای نمونه، کیشی و شیری، در مطالعات انجام شده نشان دادند که ترکیب موجک و ماشین بردار پشتیبان در پیش‌بینی دبی جریان ماهانه به نتایج بالاتری نسبت به ماشین بردار پشتیبان منجر می‌شود (Kisi and 2011). Shiri, کیشی و سیمن، توانایی ترکیب موجک با روش برنامه‌ریزی ژنتیک را در مقایسه با مدل برنامه‌ریزی ژنتیک به اثبات رساندند (Kisi and Cimen, 2011). داننده مهر و همکاران، دو مدل برنامه نویسی ژنتیکی خطی (LGP) و مدل ترکیبی موجک شبکه عصبی مصنوعی را برای پیش‌بینی جریان ماهانه دو ایستگاه پی در پی مورد مطالعه قرار دادند و به این نتیجه رسیدند که بر

جدول ۱. مشخصات جغرافیایی ایستگاه‌های هیدرومتریک متوالی رودخانه هوستونیک

Hydrometric Station	Station No.	Area(m <sup>2</sup> )	Max. discharge(m <sup>3</sup> /s)	Min. discharge(m <sup>3</sup> /s)	Max. stage(m)	Min. stage(m)
Station 1	۰۱۱۹۷۵۰۰	۷۳۰۰۰۰	۱۹۳/۴۰	۱/۷۸	۳/۶۴	۰/۶۷
Station 2	۰۱۲۰۰۵۰۰	۱۶۴۲۰۰۰	۴۹۲	۳/۵۹	۳/۴۳	۰/۳۳
Station 3	۰۱۲۰۵۵۰۰	۲۵۷۹۰۰۰	۹۲۸	۲/۶۰	۴/۹۵	۰/۳۰



شکل ۱. موقعیت ایستگاه‌های متوالی رودخانه هوستونیک

۲. در هر نقطه میانگین پوش برازش داده شده بر نقاط بیشینه محلی و پوش برازش داده شده بر نقاط کمینه محلی باید صفر باشد.

مرحله اول: تعیین نقاط بیشینه و کمینه محلی سیگنال  $x(n)$ .

مرحله دوم: به دست آوردن پوش بالایی و پایینی سیگنال با استفاده از برازش نقاط بیشینه و کمینه محلی به روش اسپلاین درجه سه (cubic spline) پس از آن محاسبه میانگین پوش بالا و پایین با نام  $m_1(n)$ .

مرحله سوم: محاسبه اختلاف میان داده و میانگین پوش بالا و پایین مطابق رابطه (۲).

چنانچه  $h_1(n)$  شرایط مربوط به یک تابع مد ذاتی را داشته باشد، درحکم اولین تابع مد ذاتی،  $IMF(n)$ ، در نظر گرفته و محاسبات به مرحله بعدی الگوریتم منتقل می‌شود. در غیر این صورت، مراحل اول تا چهارم دوباره تکرار می‌شود؛ با این تفاوت که الگوریتم به جای سیگنال اولیه  $x(n)$  روی  $h_1(n)$  اعمال می‌شود.

### روش تجزیه مد تجربی (EMD)

EMD روشی برای تجزیه سیگنال‌های گوناگون است که این عمل در فرآیندی تحت عنوان غربال کردن صورت می‌گیرد. طی این فرایند سیگنال اصلی به تعدادی مؤلفه با محتوای بسامدی متفاوت تجزیه می‌شود. روش EMD طبق رابطه (۱) سیگنال اصلی  $x(n)$  را به تعدادی مد ذاتی تجزیه می‌کند (Wu and Huang, 2004; Amirat et al., 2018).

$$x(n) = \sum_{i=1}^n c_i(x) + r_n(x) \quad \text{(رابطه ۱)}$$

که  $m(x)$  همان مولفه باقی مانده، بعد از  $n$  تعداد IMFها و  $c_i(x)$  است.

یک داده ممکن است در یک زمان دارای چندین مد ذاتی باشد. این مدهای نوسانی، تابع‌های مد ذاتی (IMF) نام دارد و دارای شرایط زیر هستند

۱. در کل داده، تعداد نقاط فرین (اکسترمم) و نقاط صفر با هم برابر و یا حداکثر دارای یک واحد اختلاف هستند.

۱. مساحت زیر منحنی آن صفر باشد.

۲. مقدار میانگین آن از هر دو سمت به صفر میل کند.

۳. ممان‌های آن هم صفر باشد، این شرط اگر وجود نداشته باشد اشکالی ندارد.

### رگرسیون فرایند گاوسی (GPR)

فرآیند گاوسی مجموعه‌ای از متغیرهای تصادفی است که تعداد محدودی از آنها با توزیع‌های گاوسی ادغام شده‌اند که تعمیمی از توزیع گاوسی می‌باشد. توزیع گاوسی در واقع توزیع بین متغیرهای تصادفی بوده در حالی که فرآیند گاوسی بیانگر توزیع بین توابع می‌باشد. فرآیند گاوسی  $f(x)$  توسط توابع میانگین  $m(x)$  و کواریانس به شکل زیر تعریف می‌شود:

$$m(x) = E(f(x)) \quad (\text{رابطه ۶})$$

(رابطه ۷)

$k(x, x') = E((f(x) - m(x))(f(x') - m(x')))$  که در روابط فوق،  $k(x, x')$  تابع کواریانس (یا کرنل) بوده که در نقاط  $x$  و  $x'$  محاسبه می‌شود. فرآیند گاوسی  $f(x)$  می‌تواند به صورت زیر بیان گردد:

$$f(x) \sim GP(m(x), k(x, x')) \quad (\text{رابطه ۸})$$

که معمولاً جهت ساده‌سازی، مقدار تابع میانگین برابر با صفر در نظر گرفته می‌شود. در فرآیند گاوسی، رابطه بین بردار ورودی و هدف به فرم زیر می‌باشد:

$$y_i = f(x_i) + \varepsilon \quad (\text{رابطه ۹})$$

که در آن  $f(x)$  بیانگر تابع رگرسیون دلخواه و  $\varepsilon$  نیز مقدار نویز توزیع گاوسی با میانگین صفر و واریانس  $\sigma^2$  می‌باشد، یعنی  $\varepsilon \sim N(0, \sigma^2)$ . علاوه بر این، چنین فرض می‌گردد که  $f = [f(x_1), f(x_2), \dots, f(x_n)]^T$  رفتاری بر مبنای فرآیند گاوسی داشته به نحوی که  $p(f|X) = N(0, K)$  که در آن  $K$  ماتریس کواریانس با درایه‌های  $k_{i,j} = k(x_i, x_j)$  می‌باشد.

(رابطه ۱۰)

$$K(X, X) = \begin{pmatrix} k(x_1, x_1) & k(x_1, x_2) & \dots & k(x_1, x_n) \\ k(x_2, x_1) & k(x_2, x_2) & \dots & k(x_2, x_n) \\ \vdots & \vdots & \ddots & \vdots \\ k(x_n, x_1) & k(x_n, x_2) & \dots & k(x_n, x_n) \end{pmatrix}$$

رگرسیون فرآیند گاوسی به منظور محاسبه توزیع پیش‌بینی شده برای مقادیر تابع  $f^*$  در نقاط تست  $X^* = [x_1^*, x_2^*, \dots, x_m^*]$  به کار می‌رود.

### معیار ارزیابی

به منظور ارزیابی کارایی روش استفاده شده برای تخمین دبی در این مقاله از سه پارامتر آماری استفاده گردیده است، که در آن

$$h_1(n) = x(n) - m_1(x) \quad (\text{رابطه ۲})$$

مرحله چهارم: محاسبه باقیمانده مطابق رابطه (۳).

$$r_1(n) = x(n) - \text{imf}_1(n) \quad (\text{رابطه ۳})$$

مرحله پنجم: چنانچه باقیمانده دارای حداقل دو فرین باشد، مراحل اول تا پنجم تکرار می‌شود. در غیر این صورت الگوریتم متوقف می‌شود و آخرین باقیمانده درحکم باقیمانده سیگنال در نظر گرفته می‌شود. پس از اینکه تجزیه داده‌های نیمرخ مغناطیسی به IMFهای مربوط صورت گرفت، باقیمانده تجزیه درحکم بی-هنجاری منطقه‌ای در نظر گرفته و اختلاف میان داده‌های مشاهده‌ای و بی‌هنجاری منطقه‌ای بی‌هنجاری باقیمانده تلقی می‌شود. با این حال، به علت وجود تناوب و نویز در سیگنال‌ها، در برخی موارد به علت اختلاط مدها، توزیع حوزه زمان-فرکانس دچار انقطاع می‌شود و عملکرد EMD دچار نقض می‌شود (چون میانگین IMFها حالتی نامشخص به خود می‌گیرد). جهت رفع این ایراد، وو و هوانگ روشی متفاوت با عنوان روش تجزیه مد تجربی یکپارچه (EEMD) را پیشنهاد کردند (Wu and 2009, Huang). در روند تجزیه‌سازی آن، حجم محدودی از نویز سفید وارد سیگنال اصلی می‌شود. با استفاده از جنبه‌های مثبت آماری نویز سفید که توزیع متوازن در حوزه فرکانس دارد، اثر نویز متناوب از روند تجزیه حذف می‌شود.

### تبدیل موجک گسسته (DWT)

موجک‌ها توابع ریاضی هستند که شکل مقیاس-زمان از سری-های زمانی و روابط آن‌ها را جهت آنالیز سری‌های زمانی که شامل متغیرها و غیرثابت‌ها می‌باشد، ارائه می‌دهد.  $\Psi(x)$  تابع موجک مادر است که توابع مورد استفاده در تحلیل، با دو عمل ریاضی انتقال و مقیاس در طول سیگنال مورد تحلیل، تغییر اندازه و تغییر محل می‌یابند.

$$\psi_{a,b}(x) = \frac{1}{\sqrt{a}} \psi\left(\frac{x-b}{a}\right) \quad (\text{رابطه ۴})$$

در نهایت ضرایب موجک در هر نقطه از سیگنال ( $b$ ) و برای هر مقدار از مقیاس ( $a$ ) با رابطه (۵) قابل محاسبه می‌باشد.

$$T(a,b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{+\infty} \psi\left(\frac{t-b}{a}\right) f(x) dt \quad (\text{رابطه ۵})$$

که در رابطه (۵)،  $a$  کار مقیاس و  $b$  کار انتقال تابع را انجام می‌دهند. به ازای مقادیر مختلف  $a$  و  $b$  مقدار  $T$  به دست می‌آید. در انتخاب موجک مادر می‌بایست شرط‌های زیر وجود داشته باشد:

شبیه‌سازی و نتایج و بحث

مدل GPR بدون تجزیه داده‌ها

جهت تخمین دبی و بررسی ارتباط بین ایستگاهی آن، مدل‌هایی بر اساس پارامترهای دبی و اشل روزانه تعریف گردیده و کارایی روش رگرسیون فرایند گاوسی مورد بررسی قرار گرفت. تمامی مدل‌ها با استفاده از داده‌های آزمایش، صحت‌سنجی می‌شوند. برای انتخاب بهترین ورودی برای گوسین، نه ترکیب ورودی انتخاب گردیده و ارزشیابی انجام گردیده است. این امر برای پیدا کردن میزان وابستگی دبی ایستگاه‌های پایین دست با ایستگاه‌های بالادست با استفاده از داده‌های بدون پیش‌پردازش انجام گردیده است. ترکیب مدل‌ها در جدول (۲) آورده شده است.

$Q_0$ : دبی اندازه‌گیری شده،  $\bar{Q}_0$ : متوسط دبی اندازه‌گیری شده،  $Q_p$ : دبی پیش‌بینی شده،  $\bar{Q}_p$ : متوسط دبی پیش‌بینی شده و  $N$  تعداد داده‌هاست. در این رابطه، DC: همبستگی خطی بین مقادیر پیش‌بینی شده و واقعی، RMSE: ریشه میانگین مربعات خطا، و R: ضریب همبستگی بین مقادیر مشاهداتی و محاسباتی می‌باشد.

$$DC = 1 - \frac{\sum_{i=1}^N (Q_0 - Q_p)^2}{\sum_{i=1}^N (Q_0 - \bar{Q}_0)^2} \quad (\text{رابطه ۱۱})$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (Q_0 - Q_p)^2}{N}} \quad (\text{رابطه ۱۲})$$

$$R = \frac{\sum_{i=1}^N (Q_0 - \bar{Q}_0) * (Q_p - \bar{Q}_p)}{\sqrt{\sum_{i=1}^N (Q_0 - \bar{Q}_0)^2 * \sum_{i=1}^N (Q_p - \bar{Q}_p)^2}} \quad (\text{رابطه ۱۳})$$

جدول ۲. ترکیب مدل‌های ورودی

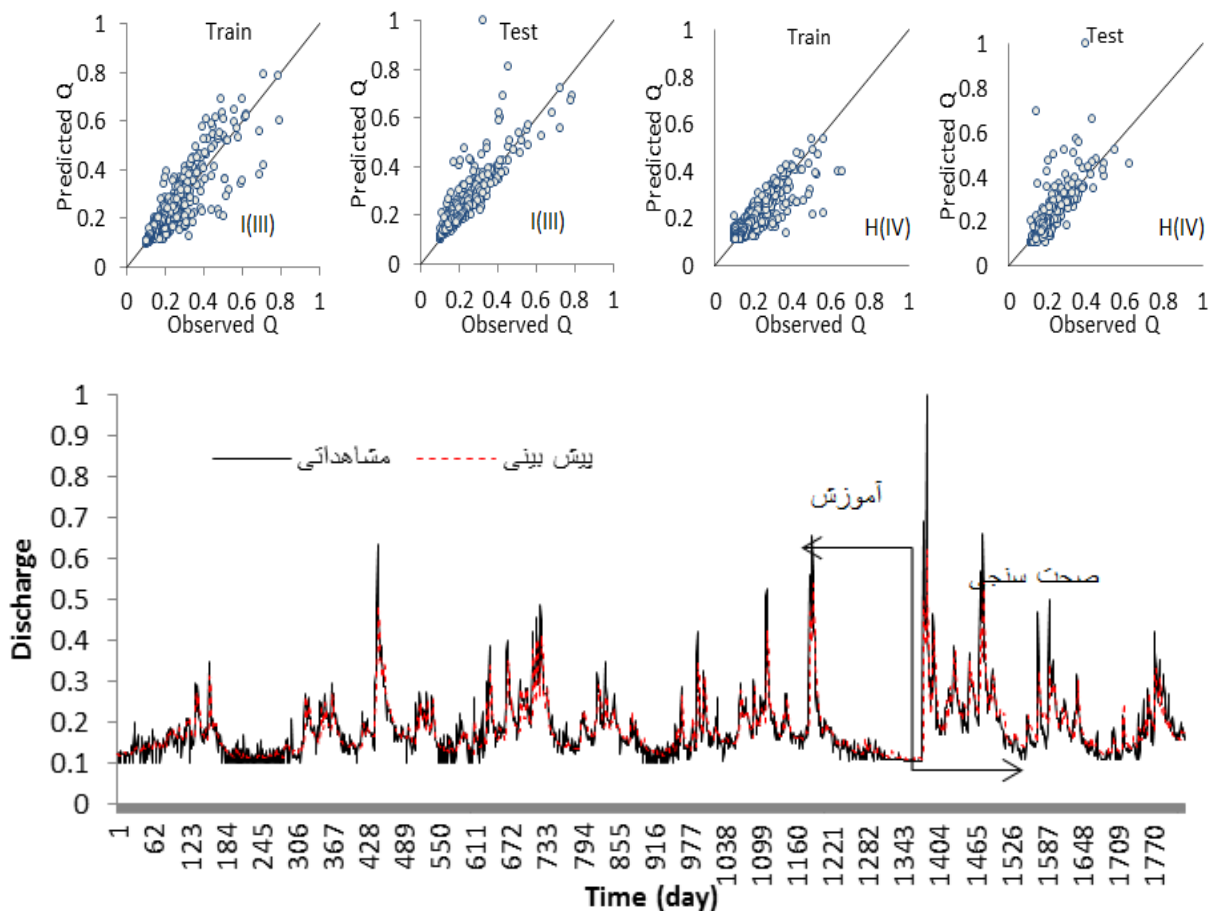
مدل‌سازی دبی ایستگاه دوم بر اساس ایستگاه اول		مدل‌سازی دبی ایستگاه سوم بر اساس ایستگاه‌های اول و دوم	
مدل	پارامترهای ورودی	مدل	پارامترهای ورودی
I(I)	$Q_2(t) = Q_1(t-1)$	H(I)	$Q_3(t) = Q_2(t-1)$
I(II)	$Q_2(t) = H_1(t-1)$	H(II)	$Q_3(t) = H_2(t-1)$
I(III)	$Q_2(t) = Q_1(t-1), H_1(t-1)$	H(III)	$Q_3(t) = Q_2(t-1), H_2(t-1)$
		H(IV)	$Q_3(t) = Q_2(t-1), Q_1(t-1)$
		H(V)	$Q_3(t) = H_2(t-1), H_1(t-1)$
		H(VI)	$Q_3(t) = Q_2(t-1), Q_1(t-1), H_2(t-1), H_1(t-1)$

برتر جهت پیش‌بینی دبی ایستگاه دوم و سوم برای سری داده‌های آموزش و آزمون و همچنین نمودار سری زمانی مرحله آموزش و آزمون مدل برتر جهت پیش‌بینی دبی ایستگاه سوم در شکل (۲) نشان داده شده است.

در این مدل‌های تعریف شده،  $Q_2(t)$  دبی ایستگاه دوم،  $Q_3(t)$  دبی ایستگاه سوم،  $Q_1(t-1)$  دبی یک روز قبل ایستگاه اول،  $H_1(t-1)$  اشل یک روز قبل ایستگاه اول،  $Q_2(t-1)$  دبی یک روز قبل ایستگاه دوم و  $H_2(t-1)$  اشل یک روز قبل ایستگاه دوم می‌باشند. نتایج معیارهای ارزیابی در جدول (۳) آورده شده است. نتایج مدل

جدول ۳. نتایج ارزیابی مدل‌ها برای روش GPR

	آموزش			آزمون		
	R	DC	RMSE	R	DC	RMSE
پیش‌بینی دبی ایستگاه دوم						
I(I)	۰/۹۰	۰/۷۵	۰/۰۴۳	۰/۸۶	۰/۷۲	۰/۰۶۰
I(II)	۰/۹۰	۰/۷۸	۰/۰۴۱	۰/۸۶	۰/۷۱	۰/۰۶۱
I(III)	۰/۹۱	۰/۷۹	۰/۰۴۰	۰/۸۷	۰/۷۴	۰/۰۵۹
پیش‌بینی دبی ایستگاه سوم						
H(I)	۰/۸۷	۰/۶۶	۰/۰۵۱	۰/۷۹	۰/۶۰	۰/۰۵۷
H(II)	۰/۸۶	۰/۶۸	۰/۰۴۹	۰/۷۷	۰/۵۸	۰/۰۶۲
H(III)	۰/۸۷	۰/۶۹	۰/۰۴۸	۰/۷۹	۰/۶۱	۰/۰۵۶
H(IV)	۰/۸۷	۰/۷	۰/۰۳۳	۰/۷۹	۰/۶۲	۰/۰۵۴
H(V)	۰/۸۷	۰/۶۵	۰/۰۵۲	۰/۷۸	۰/۶۰	۰/۰۶۰
H(VI)	۰/۸۷	۰/۶۸	۰/۰۴۹	۰/۷۹	۰/۶۰	۰/۰۵۷



شکل ۲. نتایج ارزیابی مرحله آموزش و آزمون مدل برتر روش GPR

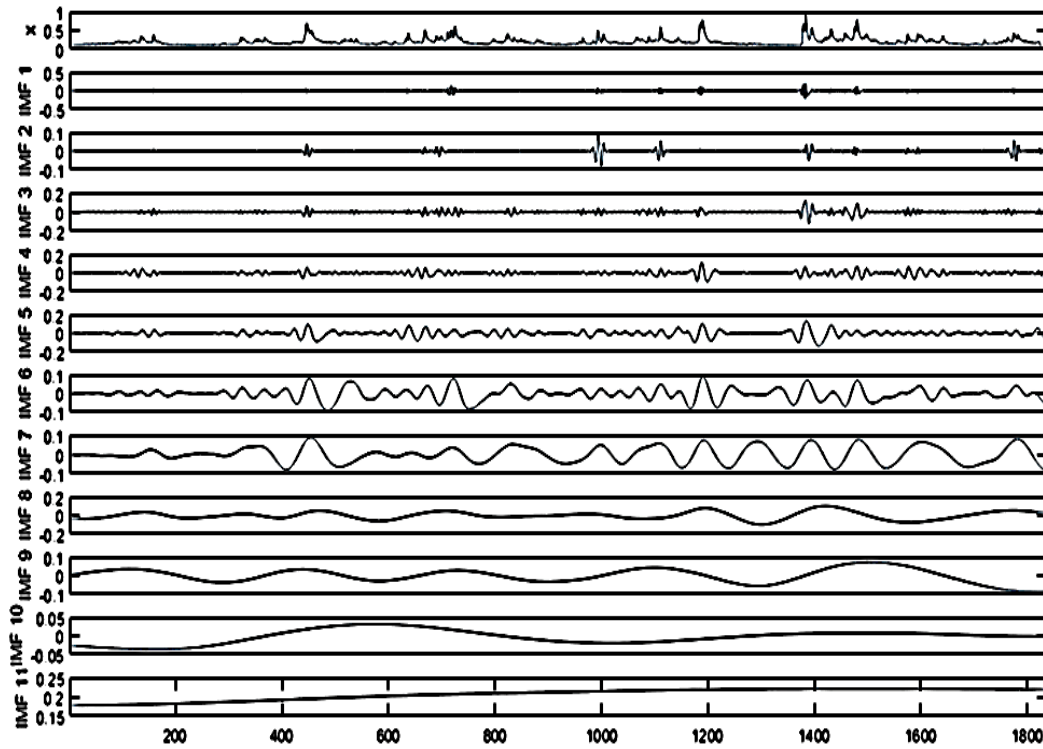
استفاده قرار گرفته است. اندازه‌ی ES برابر با ۵۰۰ برابر هر عضو، گروه EEMD انتخاب شده است (انتخاب از طریق سعی و خطا و مقایسه با مقادیر پیشنهادی استفاده شده در منابع). سری‌های زمانی طی سال‌های ۲۰۰۲-۲۰۰۶ به یازده زیرسری (IMF و Residual) تجزیه شدند. IMF‌های ۱ تا ۱۰ زیرسری‌هایی با فرکانس‌هایی از زیاد به کم هستند که هر کدام نشانگر اجزایی نوسانگر با پریودهایی مشخص می‌باشد و زیرسری باقی‌مانده (residual) عضو نشان‌دهنده روند می‌باشد که بعد از IMF10 به دست آمده است. روند محاسبه شده از طریق EEMD طی زمان تغییر می‌کند ولی مقادیر روند از گذشته بر روی سری زمانی اثرگذار نخواهد بود (در صورتی که سری زمانی از لحاظ زمانی گسترده شود). به عنوان مثال پیش‌پردازش سری‌های زمانی در مقیاس روزانه (پیش‌بینی دبی ایستگاه دوم) در شکل ۳ نشان داده شده است. در این تحقیق با استفاده از روش تجزیه مد تجزیه یکپارچه ابتدا سری زمانی اصلی به زیرسری‌هایی تجزیه گردید؛ و سپس زیر سری‌های به دست آمده به عنوان ورودی به مدل GPR

برای پیش‌بینی دبی روزانه نتایج مندرج در جدول (۲) نشان می‌دهد که بهترین مدل جهت برآورد مقدار دبی ایستگاه دوم، مدل سه با پارامتر ورودی دبی و اشل یک روز قبل ایستگاه اول و جهت برآورد دبی ایستگاه سوم، مدل چهار با پارامترهای ورودی دبی یک روز قبل ایستگاه اول و دوم می‌باشد. بهترین نتایج برای داده‌های آزمون برای پیش‌بینی دبی ایستگاه دوم با ضریب همبستگی ۰/۷۹ و جذر میانگین مربعات خطای ۰/۰۴ و برای پیش‌بینی دبی ایستگاه سوم با ضریب همبستگی ۰/۷۶ و جذر میانگین مربعات خطای ۰/۰۵۴ می‌باشد.

**مدل GPR با اعمال تجزیه مد تجزیه عمومی (EEMD-GPR)**  
 استفاده از الگوریتم غیرخطی همچون EEMD در برآورد پدیده های غیرخطی هیدرولوژیکی (در صورت وجود) می‌تواند مفید واقع شود. الگوریتم EEMD تحلیل داده‌ها را به کمک نوفه انجام می‌دهد. به بیانی دیگر، در روند تفکیک و تجزیه کردن به وسیله EEMD، نوفه سفید به داده‌های اصلی اضافی می‌شود. در تحقیق حاضر، نوفه سفید با دامنه ۰/۱ انحراف سری زمانی دبی مورد

سری زمانی مرحله آموزش و آزمون مدل برتر جهت پیش‌بینی دبی ایستگاه سوم در شکل (۴) نشان داده شده است.

داده شد. نتایج حاصل از مدل‌سازی در نظر گرفته شده در جدول (۴) ارائه شده است. نتایج مدل برتر جهت پیش‌بینی دبی ایستگاه دوم و سوم برای سری داده‌های آموزش و آزمون و همچنین نمودار

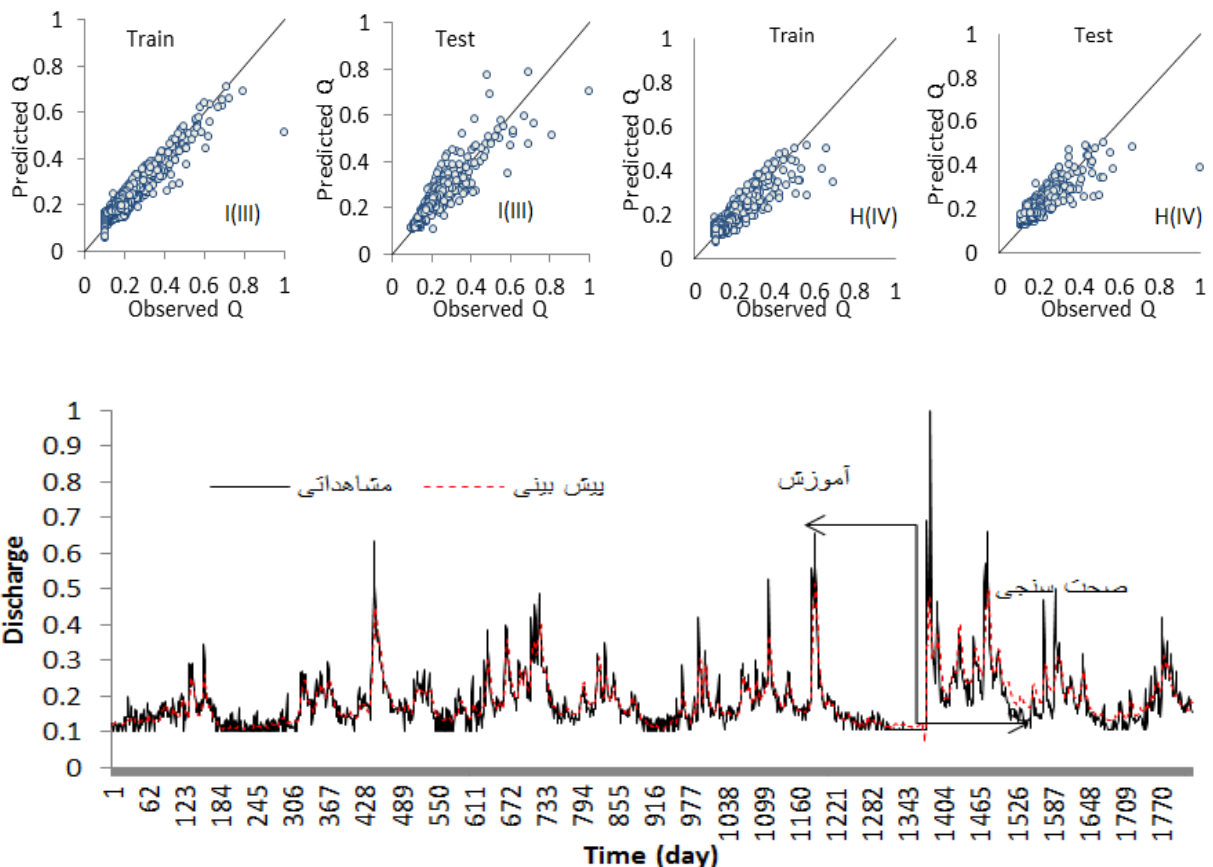


شکل ۳. زیرسری تجزیه شده توسط تجزیه مد تجزیه یکنپارچه

جدول ۴. نتایج ارزیابی مدل‌ها برای روش GPR با اعمال تجزیه مد تجزیه عمومی

مدل	آموزش			آزمون		
	R	DC	RMSE	R	DC	RMSE
پیش‌بینی دبی ایستگاه دوم						
I(I)	۰/۹۲	۰/۸۰	۰/۰۴۲	۰/۸۸	۰/۷۸	۰/۰۵۴
I(II)	۰/۹۳	۰/۸۲	۰/۰۴۰	۰/۸۸	۰/۷۷	۰/۰۵۶
I(III)	۰/۹۴	۰/۸۴	۰/۰۳۸	۰/۸۹	۰/۸۰	۰/۰۵۰
پیش‌بینی دبی ایستگاه سوم						
H(I)	۰/۹۰	۰/۷۴	۰/۰۴۹	۰/۸۲	۰/۶۸	۰/۰۵۲
H(II)	۰/۸۸	۰/۷۶	۰/۰۴۳	۰/۸۱	۰/۶۶	۰/۰۵۶
H(III)	۰/۸۹	۰/۷۷	۰/۰۴۱	۰/۸۱	۰/۶۹	۰/۰۵۱
H(IV)	۰/۹۰	۰/۸۱	۰/۰۳۱	۰/۸۵	۰/۷۰	۰/۰۵۰
H(V)	۰/۸۸	۰/۷۵	۰/۰۴۵	۰/۸۱	۰/۶۷	۰/۰۵۵
H(VI)	۰/۹۱	۰/۷۸	۰/۰۴۰	۰/۸۱	۰/۶۶	۰/۰۵۶





شکل ۴. نتایج مرحله آموزش و آزمون مدل برتر برای حالت تجزیه مد تجزیه عمومی

یک روز قبل ایستگاه اول و دوم، جهت برآورد دبی ایستگاه سوم مدل برتر می‌باشند. نمودار مدل برتر برای سری داده‌های آزمون، جهت پیش‌بینی دبی ایستگاه دوم و سوم در شکل (۴) نشان داده شده است. همچنین تأثیر هر یک از زیرسری‌های تجزیه مد تجزیه یکپارچه در بهبود نتایج بررسی گردید. مشاهده گردید که زیرسری باقیمانده، ناکارآمدترین زیرسری می‌باشد. برای نمونه نتایج مدل برتر در پیش‌بینی دبی ایستگاه دوم یکبار با تمام ۱۱ زیرسری (  $IMF_1, IMF_2, \dots, IMF_{10}, Residual$  ) و یکبار هم بدون باقیمانده (  $IMF_1, IMF_2, \dots, IMF_{10}$  ) در جدول (۵) آورده شده است.

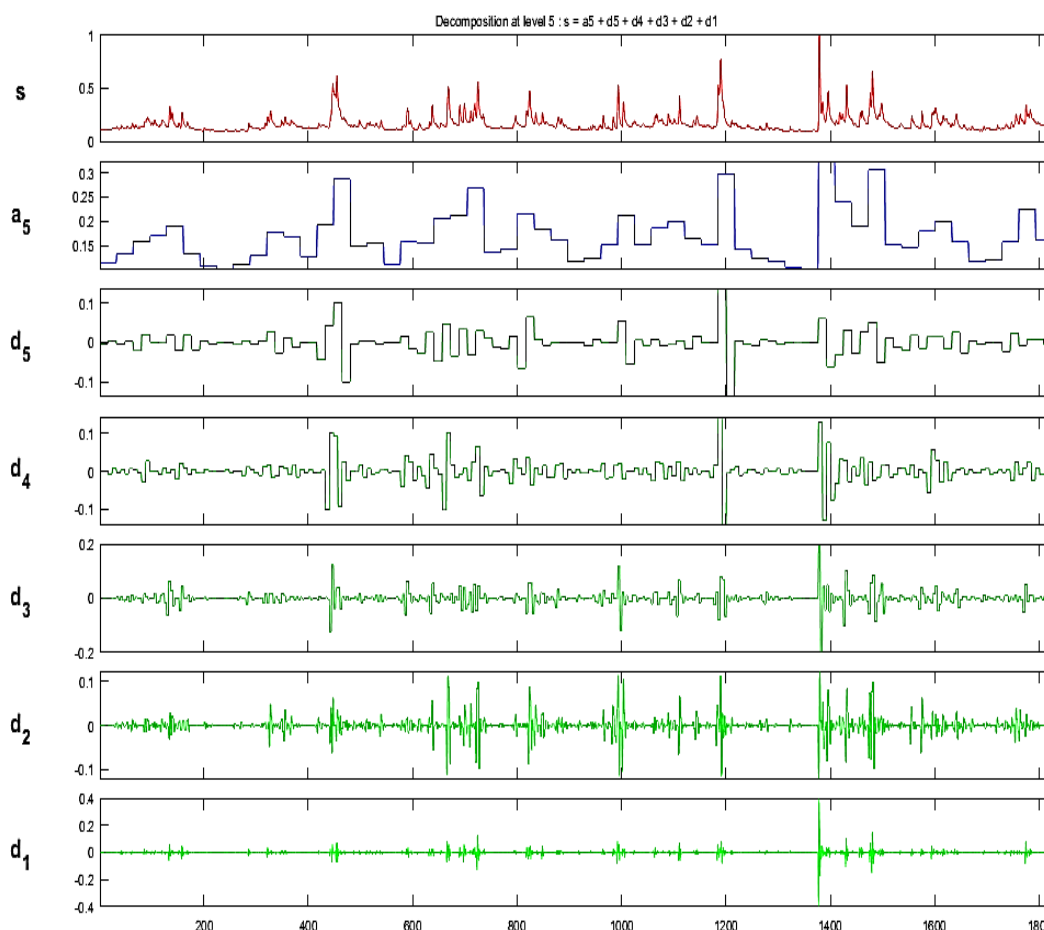
با مقایسه نتایج مربوط به روش رگرسیون فرایند گاوسی و ترکیب روش رگرسیون فرایند گاوسی-تجزیه مد تجزیه عمومی (EEMD-GPR) مشخص گردید که مدل ترکیبی ذکر شده دارای عملکرد بهتری بوده و از نظر بررسی تمام معیارهای ارزیابی نسبت به مدل گاوسین دارای عملکرد بهتری است و تجزیه زیرسری‌های زمانی تا حدود زیادی باعث بهبود نتایج گردیده است. اعمال تجزیه مد تجزیه عمومی، ضریب همبستگی را به عنوان نمونه از  $DC=0.74$  به  $DC=0.80$  ارتقاء داد و جذر میانگین مربعات خطا را کاهش داد. مطابق با جدول (۴) مدل (III) با پارامترهای ورودی دبی و اشل یک روز قبل با داشتن  $CC$  و  $DC$  بیشتر و  $RMSE$  کمتر جهت برآورد دبی ایستگاه دوم و مدل (IV) با ورودی دبی

جدول ۵. نتایج معیارهای ارزیابی مدل برتر الف) با اعمال تمام زیرسری‌های تجزیه شده (ب) با حذف باقیمانده

مدل	معیارهای ارزیابی					
	آموزش			آزمون		
	R	DC	RMSE	R	DC	RMSE
الف	۰/۹۴	۰/۸۴	۰/۰۳۸	۰/۸۹	۰/۸۰	۰/۰۵۰
ب	۰/۹۶	۰/۸۷	۰/۰۳۰	۰/۹۱	۰/۸۵	۰/۰۳۴

ورودی در مدل‌های هوش مصنوعی استفاده می‌گردد. گزینش موجک مادر منطقی، چالشی مهم در مسئله مدل ترکیبی هوش مصنوعی موجکی است که می‌تواند تأثیر قابل توجهی در نتایج مدل‌سازی داشته باشند. بدین منظور موجک خانواده دوبچی (Haar) به عنوان موجک مادر در تحقیق جاری استفاده گردید. نتایج نشان می‌دهد که براساس معیار سنجه DC، موجک مادر Haar با سطح تفکیک‌پذیری پنج عملکرد بهتری نسبت به موجک‌های دیگر در سطوح متفاوت دارد. به عنوان مثال پیش‌پردازش سری‌های زمانی در مقیاس روزانه (پیش‌بینی دبی ایستگاه دوم)، توسط تبدیل موجک گسسته در شکل ۵ نشان داده شده است.

**مدل GPR با اعمال تبدیل موجک گسسته (DWT-GPR)**  
تبدیل موجک یک روش بسیار دقیق برای پیش‌پردازش سری‌های زمانی محسوب می‌شود (Roushangar *et al.*, 2018; Aussem *et al.* 1998; Farajzadeh and Alizadeh, 2018). بنابراین در این مطالعه مدل ترکیبی موجک-گوسین برای ترکیب همزمان توانایی‌های مدل‌های هوش مصنوعی و آنالیز موجک در یک ساختار واحد برای پیش‌بینی دبی طراحی گردیده است. در این مدل، پیش‌پردازش داده‌ها توسط تبدیل موجک گسسته انجام گرفته است؛ سپس نتایج این پردازش به عنوان



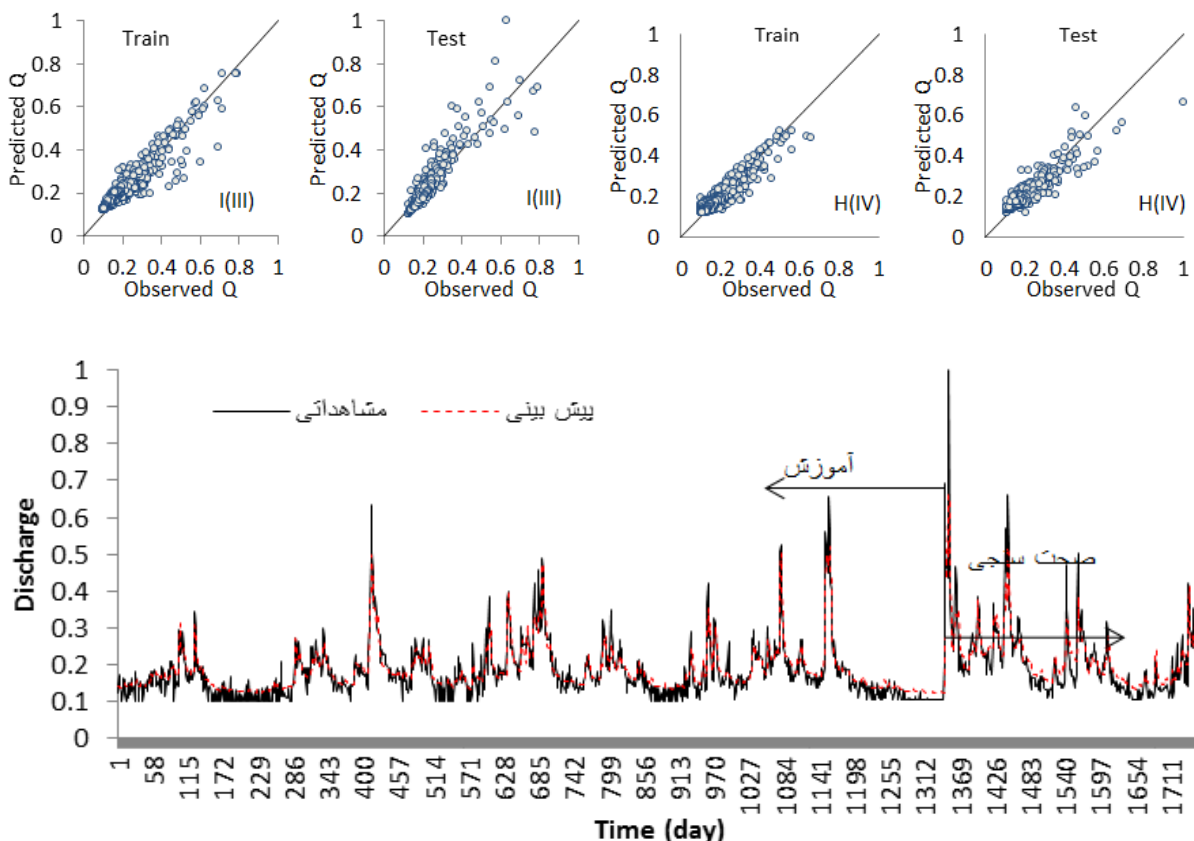
شکل ۵. زیرسری تجزیه شده توسط تبدیل موجک گسسته

برای مدل برتر جهت پیش‌بینی دبی ایستگاه دوم ارتقاء داد. نتایج حاکی از آن است که ریشه میانگین مربعات خطای روش ترکیبی موجک (DWT-GPR) و ترکیبی مد تجربی (EEMD-GPR) کمتر از روش گوسین ساده (GPR) می‌باشد. هر دو روش ترکیبی مذکور، دقت و کارایی نتایج پیش‌بینی با مدل گوسین ساده را تا حدود زیادی بهبود می‌دهند.

نتایج مربوط به اعمال تبدیل موجک در جدول (۶) آورده شده است. نتایج مدل برتر جهت پیش‌بینی دبی ایستگاه دوم و سوم برای سری داده‌های آموزش و آزمون و همچنین نمودار سری زمانی مرحله آموزش و آزمون مدل برتر جهت پیش‌بینی دبی ایستگاه سوم در شکل (۶) نشان داده شده است. با بررسی نتایج مشخص شد که مدل ترکیبی موجک عملکرد بهتری نسبت به مدل ساده گوسین دارد. برای مثال  $DC=0/74$  را به  $DC=0/83$

جدول ۶. نتایج ارزیابی مدل‌ها برای روش GPR با اعمال تبدیل موجک گسسته

مدل	آموزش			آزمون		
	R	DC	RMSE	R	DC	RMSE
پیش‌بینی دبی ایستگاه دوم						
I(I)	۰/۹۳	۰/۸۱	۰/۰۴۱	۰/۸۹	۰/۸۰	۰/۰۵۰
I(II)	۰/۹۴	۰/۸۳	۰/۰۳۹	۰/۸۹	۰/۷۹	۰/۰۵۲
I(III)	۰/۹۴	۰/۸۶	۰/۰۳۱	۰/۹۱	۰/۸۳	۰/۰۴۷
پیش‌بینی دبی ایستگاه سوم						
H(I)	۰/۹۱	۰/۷۹	۰/۰۴۵	۰/۸۳	۰/۷۷	۰/۰۴۶
H(II)	۰/۸۹	۰/۸۱	۰/۰۴۱	۰/۸۳	۰/۷۴	۰/۰۵۰
H(III)	۰/۹۰	۰/۸۲	۰/۰۳۹	۰/۸۲	۰/۷۶	۰/۰۴۸
H(IV)	۰/۹۴	۰/۸۷	۰/۰۲۹	۰/۹۱	۰/۷۹	۰/۰۳۰
H(V)	۰/۹۱	۰/۸۰	۰/۰۴۲	۰/۸۳	۰/۷۲	۰/۰۵۲
H(VI)	۰/۹۲	۰/۸۴	۰/۰۳۶	۰/۸۳	۰/۷۱	۰/۰۵۴



شکل ۶. نتایج ارزیابی مرحله آموزش و آزمون مدل برتر با روش DWT-GPR

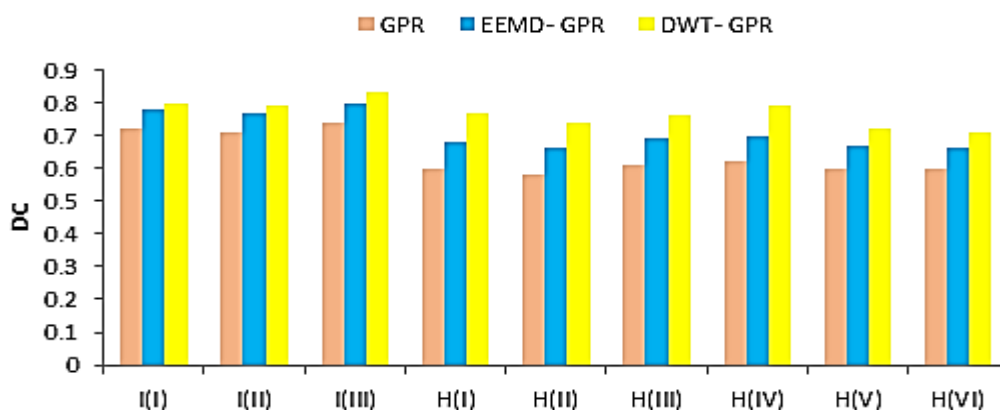
گوسین و تجزیه مد تجربی یکپارچه- گوسین (EEMD-GPR) به پیش‌بینی دبی بین ایستگاهی پرداخته شد و دبی ایستگاه پایین- دست با استفاده از داده‌های دبی ایستگاه بالادست تخمین زده

### مقایسه نتایج

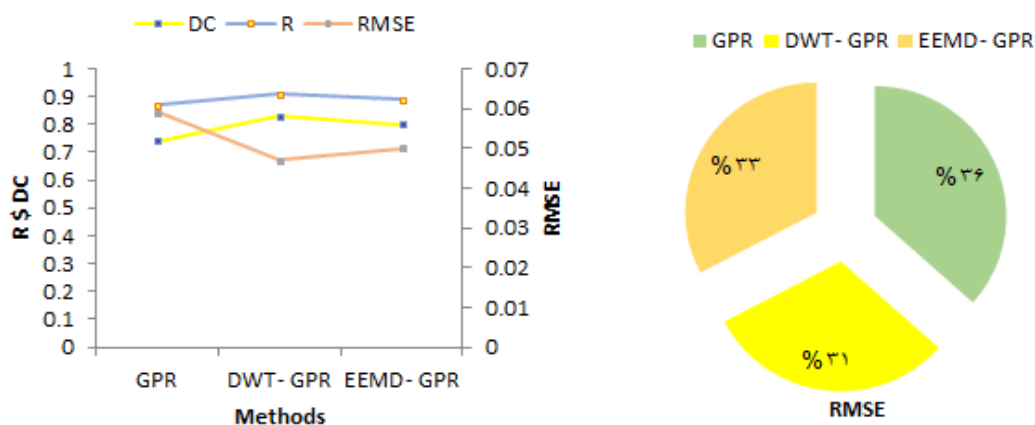
در این مطالعه با استفاده از روش‌های مبتنی بر هوش مصنوعی مانند مدل رگرسیون فرایند گاوسی و مدل‌های ترکیبی موجک-

عملکرد بهتری دارند که به عنوان نمونه روش ترکیبی موجک-گوسین برای مدل برتر مرحله آزمون  $DC=0/74$  را به  $DC=0/83$  ارتقاء داد و  $RMSE=0/059$  را به  $RMSE=0/047$  کاهش داد. همچنین مدل ترکیبی تجزیه مد تجربی یکپارچه-گوسین برای مرحله آزمون مدل برتر  $DC=0/74$  را به  $DC=0/80$  ارتقاء داد و  $RMSE=0/059$  را به  $RMSE=0/050$  کاهش داد. مقایسه نتایج معیارهای ارزیابی هر سه روش، برای مدل برتر در پیش‌بینی دبی ایستگاه دوم در شکل (۸) آورده شده است.

شد. نتایج معیارهای ارزیابی برای مرحله آزمون تمامی مدل‌ها با استفاده از روش‌های مذکور در شکل (۷) آورده شده است. مطابق نتایج شکل (۷) مشاهده گردید که مدل (III) با پارامترهای ورودی دبی و اشل یک روز قبل ایستگاه اول در تخمین دبی ایستگاه دوم و مدل (IV) در تخمین دبی ایستگاه سوم با پارامترهای ورودی دبی یک روز قبل ایستگاه اول و دوم با داشتن  $DC$  بیشتر و  $RMSE$  کمتر نسبت به سایر مدل‌ها، مدل‌های برتر می‌باشند. روش‌های ترکیبی مذکور نسبت به مدل گوسین ساده



شکل ۷. مقایسه ضریب تبیین برای مرحله آزمون تمام مدل‌ها با استفاده از سه مدل (گوسین، موجک-گوسین و تجزیه مد تجربی یکپارچه-گوسین)



شکل ۸. مقایسه معیارهای ارزیابی مدل برتر جهت پیش‌بینی دبی ایستگاه دوم برای مرحله آزمون با استفاده از سه مدل (گوسین، موجک-گوسین و تجزیه مد تجربی یکپارچه-گوسین)

ایستگاهی رودخانه هوستونیک در مقیاس زمانی روزانه استفاده شده است. به منظور ارتقاء کارایی مدل و در نظر گرفتن اثرات تبدیل موجک و تجزیه مد تجربی یکپارچه که دارای توانایی‌های تشخیص ویژگی‌های چندگانه مقیاس سری‌های زمانی می‌باشند و اطلاعات پنهان در سری زمانی را استخراج می‌کنند؛ جهت تفکیک سری‌های زمانی دبی-اشل به زیرسری‌های متفاوت استفاده شده‌اند. سپس زیرسری‌ها به عنوان ورودی برای پیش-بینی دبی در مدل گوسین استفاده شده‌اند و اثر تبدیل موجک و

## نتیجه گیری کلی

برای کاربردهای مختلف هیدرولوژیکی، مهندسی رودخانه و کنترل پروژه‌های منابع آبی، اطلاعات دقیق درباره میزان جریان رودخانه‌ها بسیار مهم است. تا کنون روابط گوناگون و پیچیده‌ای به منظور پیش‌بینی دبی رودخانه ارائه شده است. با توجه به اهمیت پیش‌بینی دبی، استفاده از مدل‌های هوشمند جعه سیاه مانند رگرسیون فرایند گوسین می‌تواند به برآورد دقیق‌تر دبی منجر گردد. مدل رگرسیون گاوسی، برای پیش‌بینی دبی بین

گردید که زیرسری باقیمانده (Residual) ناکارآمدترین زیرسری است. با توجه به نتایج معیارهای ارزیابی، مدل پیشنهادی در این تحقیق، مدل ترکیبی موجک-گوسین می‌باشد. همچنین بیشترین خطا در مدل‌سازی زمانی رخ داد که پارامتر اشل یک روز قبل به تنهایی به عنوان ورودی در نظر گرفته شد؛ و مدل I(III) با ورودی دبی و اشل یک روز قبل از ایستگاه اول، جهت تخمین دبی ایستگاه دوم و مدل H(IV) با پارامترهای ورودی دبی یک روز قبل از ایستگاه اول و دوم، جهت تخمین دبی ایستگاه سوم با داشتن خطای مدل‌سازی کمتر، مدل‌های برتر در این پژوهش می‌باشند.

## REFERENCES

- Adarsh, S. VishnuPriya, M. S. Narayanan, S. Smruthi, M. S. George, P. & Benjie, N. M. (2016), Trend analysis of sediment flux time series from tropical river basins in India using non-parametric tests and multiscale decomposition. *Modeling Earth Systems and Environment*, 2(4), 187.
- Adnan, R. M. Yuan, X. Kisi, O. & Yuan, Y. (2017). Streamflow Forecasting Using Artificial Neural Network and Support Vector Machine Models. *American Scientific Research Journal for Engineering, Technology, and Sciences (ASRJETS)*, 29(1), 286-294.
- Alizadeh, M. J. Kavianpour, M. R. Kisi, O. & Nourani, V. (2017). A new approach for simulating and forecasting the rainfall-runoff process within the next two months. *Journal of Hydrology*, 548, 588–597.
- Amirat, Y. Benbouzidb, MEH. Wang, T. Bacha, K and Feld, G. (2018), EEMD-based notch filter for induction machine bearing faults detection, *Applied Acoustics*, 133: 202–209.
- Aussem, A. Campbell, J. and Murtagh, F. (1998), Wavelet-based feature extraction and decomposition strategies for financial forecasting, *Journal of Computational Finance*, 6 (2): 5–12.
- Behzadi, M. Asghari, K. Aazi, M and Palhang, M. (2009), Generalization performance of support vector machines and neural networks in runoff modeling, *Expert Systems with Applications*, 36: 7624-7629.
- Danandeh Mehr, A. Kahya, E and Olyaie, E.(2013), Streamflow prediction using linear genetic programming in comparison with a neuro-wavelet technique, *Journal of Hydrology*, 505: 240-249.
- Farajzadeh, J and Alizadeh, F. (2018), A hybrid linear-nonlinear approach to predict the monthly rainfall over the Urmia Lake watershed using wavelet-SARIMAX-LSSVM conjugated model, *Journal of Hydroinformatics*, 20 (1): 246–262.
- Kisi, O. and Cobaner, M. (2009), Modeling river stage-discharge relationship using different neural network computing techniques, 37 (2): 160-169.
- Kisi, O. and Cimen, M. (2011), A wavelet-support vector machine conjunction model for monthly stream flow forecasting, *Journal of Hydrology*, 399: 132–140.
- Kisi O. and Shiri J. (2011), Precipitation forecasting using wavelet-genetic programming and wavelet-neuro-fuzzy conjunction models, *Water Resource management*, 25: 3135–3152.
- Modaresi, F. Araghinejad, S. & Ebrahimi, K. (2017). A Comparative Assessment of Artificial Neural Network, Generalized Regression Neural Network, Least-Square Support Vector Regression, and K-Nearest Neighbor Regression for Monthly Streamflow Forecasting in Linear and Nonlinear Conditions. *Water Resources Management*, 1-16.
- Roushangar, K. Mehrabani, F. V. and Alami, M. (2013), Forecasting daily stream flows of vaniar river using Genetic Programming and Neural Networks approaches, *J. Civil Eng, Urban*, 3 (4): 197- 200.
- Roushangar, K. Alizadeh, F. and Adamowski, J. (2018), Exploring the effects of climatic variables on monthly precipitation variation using a continuous wavelet-based multiscale entropy approach, *Environmental Research*, 165: 176–192.
- Seyam, M. Othman, F. & El-Shafie, A. (2017). Prediction of Stream Flow in Humid Tropical Rivers by Support Vector Machines. In *MATEC Web of Conferences* (Vol. 111, p. 01007). EDP Sciences.
- Tiwari, M. K. & Adamowski, J. F. (2014), Medium-term urban water demand forecasting with limited data using an ensemble wavelet-bootstrap machine-learning approach. *Journal of Water Resources Planning and Management*, 141(2): 04014053.
- Wu, Z. Huang, NE. (2004), A study of the characteristics of white noise using the empirical mode decomposition method, *Proc RS Lond* 460A: 1597–1611.
- Wu, Z. and Huang, N. E. (2009), Ensemble empirical mode decomposition: a noise-assisted data analysis method, *Advances in Adaptive Data Analysis*, 104 (38): 14889–14894.