

مقاله علمی-پژوهشی

پیش‌بینی مقادیر TDS و EC رودخانه با استفاده از روش‌های یادگیری ماشین

محمدهادی جبار مطوری^۱، احمد فتحی^۲، فرشاد احمدی^{۳*}

تاریخ دریافت: ۱۴۰۱/۰۶/۰۳ تاریخ پذیرش: ۱۴۰۱/۰۶/۲۷

چکیده

در این مطالعه ضمن بررسی عملکرد دو مدل جنگل تصادفی و رگرسیون فرایند گاوسی، به شبیه‌سازی مقادیر هدایت الکتریکی و کل مواد جامد محلول رودخانه در ایستگاه‌های غرب دریاچه ارومیه (چهریق علیا، دیزج و تپیک) با توجه به مقادیر دبی و بی‌کربنات‌ها در دوره آماری ۹۹-۱۳۵۰ پرداخته شده است. میزان خطای شبیه‌سازی مقادیر EC با استفاده از مدل RF نسبت به مدل GPR در دو فاز آزمایش در ایستگاه‌های چهریق علیا، دیزج و تپیک حدود ۳۵۶، ۳۶ و ۴۷ درصد کمتر می‌باشد. در مورد شبیه‌سازی مقادیر TDS در ایستگاه‌های مورد مطالعه نیز، در ایستگاه چهریق علیا مدل RF و در دو ایستگاه دیگر، مدل GPR میزان خطای کمتر و کارایی بالاتری ارائه کرد. در مورد دو ایستگاه دیزج و تپیک در فاز آموزش تفاوت معنی‌داری بین میزان خطای شبیه‌سازی دو مدل RF و GPR مشاهده نشد، اما در فاز آزمایش، میزان خطای مدل GPR نسبت به مدل RF به ترتیب حدود ۶۶ و ۷۴ درصد کمتر می‌باشد. به‌طور کلی نتایج نشان داد که با توجه به حدود اطمینان شبیه‌سازی شده پارامترهای EC و TDS، کارایی دو مدل مورد بررسی قابل قبول می‌باشد، اما در مورد مقادیر TDS، رفتار دو مدل مورد بررسی در دو ایستگاه دیزج و تپیک متفاوت می‌باشد.

واژه‌های کلیدی: جنگل تصادفی، دریاچه ارومیه، رگرسیون، فرایند گاوسی، کیفیت آب

مقدمه

بهترین وجه تعداد زیادی از اهداف شناسایی شده را برآورده می‌کند (Palani et al., 2008). از طرف دیگر کمبود آب یک تهدید رو به رشد برای توسعه اقتصادی و اجتماعی است و آلودگی گسترده آب در دهه‌های اخیر این تهدید را پیچیده‌تر کرده است (Duan et al., 2016). پیش‌بینی کیفیت آب به‌عنوان یک ابزار مهم برای جلوگیری از آلودگی آب، نقش مهمی در مدیریت منابع آب ایفا می‌کند. تحقیقات نشان داده است که برخی از آلاینده‌ها ممکن است برای مسافت‌های طولانی توسط رودخانه‌ها منتقل شوند (Chau, 2005). بنابراین، به دست آوردن وضعیت تغییرات کیفیت آب از قبل و اتخاذ تدابیری برای جلوگیری از آلودگی گسترده‌تر حیاتی است. علاوه بر این، بیان دقیق کیفیت آب در یک منطقه خاص شواهد علمی برای توسعه اقتصادی، برنامه‌ریزی تجاری و حفاظت از منابع آب در برابر آلودگی‌های آینده ارائه می‌کند (Wang et al., 2014). تا به امروز، روش‌های زیادی برای شبیه‌سازی و مدل‌سازی پارامترهای مختلف هواشناسی و هیدرولوژیک با استفاده از تکنیک‌های مختلف مانند روش‌های هوشمند شبکه عصبی، برنامه‌ریزی ژنتیک، روش‌های مبتنی بر رگرسیون، مدل‌های خطی و غیرخطی و غیره استفاده شده و مورد بررسی قرار گرفته است. روش‌های رگرسیون به دلیل انعطاف‌پذیری و صریح بودن، توجه بیشتری را به خود جلب کردند. رویکردهای متنوع

جریان رودخانه به‌عنوان منبع مهمی برای آب آشامیدنی و تغذیه آب زیرزمینی عمل می‌کند (Adamowski and Sun, 2010; Emamgholizadeh et al., 2014). با توجه به وابستگی شدید کیفیت آب‌های سطحی (یعنی ویژگی‌های شیمیایی، فیزیکی و بیولوژیکی آن) به ماهیت و وسعت فعالیت‌های کشاورزی، صنعتی و سایر فعالیت‌های انسانی در حوضه‌های آبریز یک منطقه، ضروری است که پیش‌بینی آن دقیق باشد (Singh et al., 2009; Najah et al., 2013). چنین دقتی کلیدی است که یک مدیر کیفیت آب رودخانه را قادر می‌سازد تا گزینه مدیریتی را انتخاب کند که به

۱- دانشجوی کارشناسی ارشد عمران- محیط زیست، دانشکده مهندسی آب و محیط زیست، دانشگاه شهید چمران اهواز، اهواز
۲- استادیار گروه مهندسی محیط زیست، دانشکده مهندسی آب و محیط زیست، دانشگاه شهید چمران اهواز، اهواز
۳- استادیار گروه هیدرولوژی و منابع آب، دانشکده مهندسی آب و محیط زیست، دانشگاه شهید چمران اهواز، اهواز
(*- نویسنده مسئول: (Email:F.ahmadi@scu.ac.ir)

القهرطانی و همکاران مقایسه مدل‌های یادگیری ماشین تحت نظارت فردی، مانند برنامه‌ریزی بیان ژن (GEP) و شبکه عصبی مصنوعی (ANN)، با مدل یادگیری گروهی، یعنی جنگل تصادفی (RF)، برای پیش‌بینی آب رودخانه را با استفاده از مقادیر هدایت الکتریکی (EC) و جامدات محلول (TDS) در حوضه رودخانه سند، پاکستان انجام دادند. بهینه‌سازی مدل RF با تولید ۲۰ مدل فرعی به منظور انتخاب دقیق به دست آمد. نکویی برازش مدل‌ها از طریق شاخص‌های آماری معروف، مانند ضریب تعیین (R^2)، میانگین خطای مطلق (MAE)، ریشه میانگین مربعات خطا (RMSE) و کارایی نش-ساتکلیف (NSE) ارزیابی شد. نتایج نشان‌دهنده ارتباط قوی بین ورودی‌ها و خروجی‌های مدل‌سازی بود که مقدار R^2 برای مدل‌های GEP، RF و ANN به ترتیب ۰/۹۶، ۰/۹۸ و ۰/۹۲ بود. مقایسه نتایج نشان داد که روش‌های پیشنهادی برتری نسبی RF را در مقایسه با GEP و ANN دارد. همچنین نتایج تجزیه و تحلیل حساسیت نشان داد که HCO_3 مؤثرترین متغیر به برای هر دو پارامتر EC و TDS است. در نهایت، نتایج تحقیق حاضر نشان داد که مدل RF با پارامترهای کلیدی انتخاب شده می‌تواند برای ارزیابی و مدیریت کیفیت آب در اولویت قرار گیرد (Alqahtani et al., 2022). وانگ و همکاران با استفاده از رگرسیون جنگل تصادفی (RFR) به مدل‌سازی کیفیت آب برای حوضه دریاچه تایهو در استان ژجیانگ چین پرداختند. نتایج تحقیقات آن‌ها نشان داد که مدل RFR یک الگوی توزیع پیوسته از کیفیت آب را ارائه می‌کند که اطلاعات مهمی را برای کمک به پروژه‌های احیای کیفیت آب ارائه می‌کند (Wang et al., 2021). مدل‌های مبتنی بر رگرسیون پایه به دلیل ایجاد وزن مناسب برای هر پارامتر و دخالت تأثیرپذیری آن بر متغیرهای مختلف، می‌تواند عملکرد مناسبی در خصوص شبیه‌سازی و پیش‌بینی ارائه کند. بر این اساس، هدف این مطالعه، شبیه‌سازی و پیش‌بینی متغیرهای کیفی رودخانه اعم از هدایت الکتریکی (EC) و کل مواد جامد محلول (TDS) متأثر از دبی جریان و بی‌کربنات‌ها با استفاده از مدل‌های جنگل تصادفی و رگرسیون گاوسی در ایستگاه‌های دیزج، چهریق علیا و تپیک به ترتیب در زیرحوضه‌های باراندوزچای، زولاچای و نازلوچای در غرب دریاچه ارومیه می‌باشد. در این مطالعه دو پارامتر EC و TDS به صورت مجزا شبیه‌سازی شده‌اند. هدف این مطالعه شبیه‌سازی این دو پارامتر به صورت مجزا و با استفاده از مقادیر دبی جریان و بی‌کربنات‌ها می‌باشد. از آنجا کاهش دبی جریان باعث افزایش غلظت املاح در رودخانه می‌شود، در شبیه‌سازی مقادیر EC و TDS پارامترهای مؤثر می‌تواند دقت شبیه‌سازی را افزایش دهد. مقادیر EC و TDS به دلیل اهمیت آن در بحث کشاورزی و شرب به عنوان پارامترهای اصلی انتخاب شده و به دلیل وجود همبستگی بین این دو متغیر و پارامترهای دبی جریان و بی‌کربنات‌ها، این دو پارامتر نیز در شبیه‌سازی‌ها دخالت داده شد. نوآوری این تحقیق نیز شبیه‌سازی سه

متعددی از تکنیک‌های شبیه‌سازی مبتنی بر رگرسیون مانند مدل رگرسیون لجستیک (Abaurrea and Asin, 2005)، رگرسیون چند جمله‌ای محلی (George et al., 2016)، رگرسیون خطی و غیرخطی (Wilby et al., 2002)، تحلیل همبستگی متعارف (Landman & Mason, 2001)، مؤلفه‌های اصلی (Chu et al., 2008) وجود دارد، ماشین بردار پشتیبان (Tripathi et al., 2006; Ghosh et al., 2010; Raje et al., 2011; Kundu et al., 2017) مورد بررسی قرار گرفته است. مطالعات متعددی وجود دارد که ثابت کرد انواع روش‌های مبتنی بر رگرسیون نسبت به ANN برتری دارد (Mendal et al., 2011; Chen et al., 2016; Yoon et al., 2011; Eslami et al., 2022)، شبیه‌سازی رواناب و رسوب (Misra et al., 2009)، پیش‌بینی سیلاب (Liong and Sivapragasam, 2002)، مدل‌سازی بارش-رواناب (Dibike et al., 2001; Tabatabaei et al., 2021)، پیش‌بینی پارامترهای تصفیه فاضلاب (Raji et al., 2022)، شبیه‌سازی دمای نقطه شبنم (Nazeri-Tahroudi and Ramezani, 2020)، پیش‌بینی جریان رودخانه (He et al., 2014; Bhagwat and Maity, 2012) و مدل‌سازی رابطه دبی-رسوب معلق (Kisi, 2012) می‌توان عملکرد بهتر روش‌های رگرسیونی را نسبت به مدل‌های هوشمند مشاهده کرد.

در مطالعه‌ای هدام و کیسی، با استفاده از یادگیری مقادیر حدی^۱ (ELM) مرسوم و برخی مدل‌های بهبودیافته، مانند ELM متوالی آنلاین^۲ (OS-ELM) و ELM مبتنی بر شبکه RBF (R-ELM) اقدام به پیش‌بینی پارامترهای هشت سایت منتخب سازمان زمین‌شناسی ایالات متحده^۳ (USGS) کردند. این تحقیق نشان داد که اثر پیش‌بینی ELM و مدل بهبود یافته آن برای سایت‌های نظارتی مختلف و پارامترهای مختلف دقیقاً یکسان نیست، به این معنی که یک مدل برای یک سایت یا یک پارامتر خوب عمل می‌کند اما برای دیگری ضعیف است. (Heddum and Kisi, 2017). تانگ و یاسین جهت محاسبه شاخص کیفیت آب مبتنی بر شش متغیر کیفیت آب (یعنی اکسیژن محلول در درصد اشباع، نیاز اکسیژن بیوشیمیایی، نیاز شیمیایی اکسیژن، pH، مواد جامد معلق و نیتروژن آمونیاکی از مدل‌های طبقه‌بندی یادگیری عمیق (DL)، جنگل تصادفی (RF) و درخت تصمیم (DT) بر اساس دو سناریو مدل‌سازی بررسی شده S-I و S-II استفاده کردند. آن‌ها در هر دو سناریو، مدل DL را به عنوان مدل برتر معرفی کردند. آن‌ها بیان کردند که مدل RF از نظر دقت عملکرد مدل و همچنین میزان خطای طبقه‌بندی نسبتاً بهتر عمل می‌کند (Tung and Yaseen, 2021)

1 - Extreme learning machine

2 - Online sequential ELM

3 - US Geological Survey

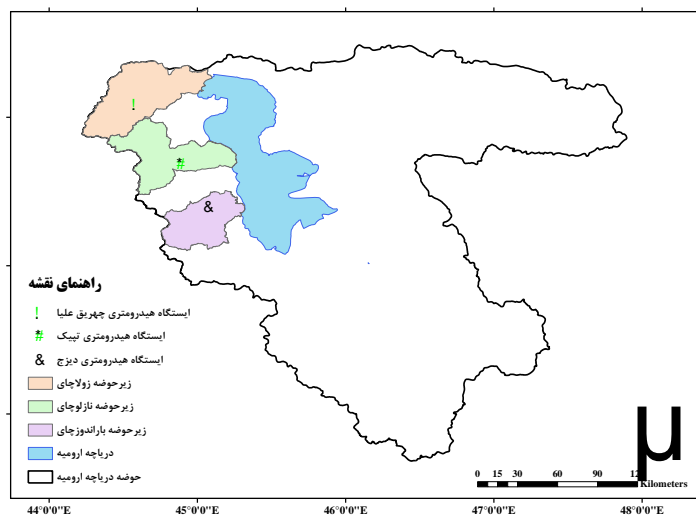
الکتریکی (EC) و کل مواد جامد محلول (TDS) در دوره آماری ۱۳۵۰-۱۳۹۹ در زیرحوضه‌های غربی دریاچه ارومیه واقع در شمال غرب ایران، شامل زیرحوضه زولاچای، نازلوچای و باراندوزچای به شرح شکل ۱ استفاده خواهد شد. از آنجایی که تغییرات دبی جریان تأثیر زیادی بر پارامترهای کیفی رودخانه دارند، لذا در این مطالعه از مقادیر دبی جریان متناظر با مقادیر کیفی نیز در مدل‌سازی و شبیه‌سازی‌ها استفاده شده است. جدول ۱ نیز مشخصات آماری داده‌های مورد بررسی را ارائه می‌دهد.

متغیره مقادیر EC و TDS به صورت مجزا به شرط وقوع مقادیر دبی جریان و بی‌کربنات‌ها می‌باشد که توسط دو مدل جنگل تصادفی و رگرسیون فرآیند گاوسی انجام شده است. از طرفی این دو مدل نیز در شبیه‌سازی مقادیر کیفی متناظر با مقادیر کمی کمتر مورد بررسی قرار گرفته است.

مواد و روش‌ها

منطقه مورد مطالعه

در این مطالعه از داده‌های بی‌کربنات‌ها (HCO_3^-)، هدایت



شکل ۱- موقعیت زیرحوضه‌های مورد مطالعه در غرب دریاچه ارومیه

جدول ۱- مشخصات آماری داده‌های مورد بررسی در دوره آماری ۱۳۵۰-۹۹

Sub-basin	Parameter	Average	Max	Min
Tapik	Discharge (m^3/s)	۱۵/۲۷	۸۴/۸۰	۳/۴۸
	TDS (mg/l)	۲۹۱/۱۹	۳۷۸/۶۴	۲۲۲/۱۷
	EC (mmho/cm)	۴۴۳/۹۸	۵۲۸/۲۷	۳۴۱/۵۰
	HCO_3^- (mg/l)	۳/۷۶	۴/۷۰	۲/۳۷
Chehriq-Olya	Discharge (m^3/s)	۴/۲۸	۱۶/۲۳	۱/۱۷
	TDS (mg/l)	۲۸۰/۰۴	۳۵۴/۰۰	۲۴۱/۰۰
	EC (mmho/cm)	۴۳۳/۰۲	۵۳۱/۰۰	۳۷۰/۲۵
	HCO_3^- (mg/l)	۳/۶۹	۴/۴۰	۲/۷۵
Dizaj	Discharge (m^3/s)	۸/۷۵	۱۸/۲۵	۳/۱۰
	TDS (mg/l)	۲۷۳/۶۳	۳۸۶/۰۰	۲۱۰/۴۰
	EC (mmho/cm)	۴۱۷/۴۳	۴۹۷/۰۸	۳۳۱/۵۴
	HCO_3^- (mg/l)	۳/۶۶	۴/۵۰	۲/۳۳

من^۱ مطرح و در سال ۱۹۷۹ توسط کندال^۲ توسعه یافت. آزمون من-

بررسی تغییرات مقادیر مشاهداتی

یکی از روش‌های مهم و پرکاربرد در زمینه بررسی روند تغییرات مقادیر مختلف، آزمون من- کندال می‌باشد که در سال ۱۹۴۵ توسط

1 - Mann

2 - Kendall

Saha et al (2020), Kim et al (2020) et al (2019a&b) (2020) Virmani et al (2021) و Iannace et al اشاره کرد. در این میان، بسیاری به مدل‌های رگرسیون عددی مربوط می‌شوند (Weisberg, 2005). در ساده‌ترین حالت مسئله رگرسیون خطی، یک متغیر وابسته y وجود دارد که فرض می‌شود این متغیر می‌تواند به‌عنوان تابعی از یک متغیر مستقل x مدل شود (Montgomery et al., 2021).

برای غلبه بر محدودیت‌های رگرسیون خطی، یک رویکرد گاوسی برای مسئله پیشنهاد شد که از نیاز به تنظیم پیش‌گیرانه پارامترها رهایی یافته و اقدام خود را در جستجوی توزیعی بر روی توابع ممکن مطابق با داده‌های مشاهده شده نهایی می‌کند. رگرسیون گاوسی بر اساس یک رویکرد بیزین با تعیین توزیع احتمال بر روی تمام مقادیر ممکن موجود به‌عنوان ورودی به سیستم عمل می‌کند (Kim and Gau, 2004). رویکرد بیزین به رگرسیون، به‌جای مقادیر دقیق برای هر پارامتر در یک تابع پیشنهادی، توزیع احتمال را بر روی تمام مقادیر ممکن بر اساس رویکرد کلاسیک تعیین می‌کند (Birgé, 2004). برای این کار، از قضیه بیزین به‌صورت زیر استفاده می‌شود:

$$p(\beta | Y, X) = \frac{p(Y | X, \beta)p(\beta)}{p(Y | X)} \quad (4)$$

که در آن $p(\beta | Y, X)$ برابر با احتمال پسین، $p(Y | X, \beta)$ برابر با تابع درست‌نمایی، $p(\beta)$ برابر با احتمال پیشین و $p(Y | X)$ برابر با احتمال حاشیه‌ای است. پروسه گاوسی با استفاده از تابع میانگین و تابع کوواریانس تعریف می‌شود. سپس، اگر $f(x)$ یک فرآیند گاوسی با تابع میانگین m و تابع کوواریانس k باشد، خواهیم نوشت:

$$f(x) \sim GP(m(x), k(x, x')) \quad (5)$$

مجموعه داده ما شامل مشاهدات X و برچسب‌های نسبی y است که به دو زیر مجموعه تقسیم می‌شوند که به ترتیب در مرحله آموزش و متعاقباً در مرحله آزمایش استفاده خواهند شد (Burt et al., 2019). فرآیند گاوسی احتمال مشترک نتایج را تعریف می‌کند، که برخی از آن‌ها از داده‌های مشاهداتی و برخی دیگر به مجموعه‌ای که قرار است آزمایش شود اشاره دارد:

$$\begin{bmatrix} y \\ f_* \end{bmatrix} \sim N \left(\begin{bmatrix} \mu \\ \mu_* \end{bmatrix}, \begin{bmatrix} K(X, X) & K(X, X_*) \\ K(X_*, X) & K(X_*, X_*) \end{bmatrix} \right) \quad (6)$$

که در آن y برابر با خروجی مرحله آموزش، f_* برابر با خروجی مرحله آزمایش و X_* نیز برابر با مجموعه آزمایش می‌باشد. K تابع کوواریانس است که تابع کرنل نیز نامیده می‌شود، و ویژگی‌های این تابع بین رویکردهای مختلفی که در آن‌ها روش مبتنی بر رگرسیون گاوسی قابل پرداخت است، تمایز قائل می‌شود (Yu et al., 2019; Sain et al., 1994). در این مطالعه از دو معیار جذر میانگین مربعات خطا (RMSE) و ضریب نش-ساتکلیف (NSE) جهت ارزیابی

کندال، آزمونی ناپارامتری بوده که در تحقیقات مختلف مورد بررسی قرار گرفته و دقت آن تأیید شده است (Khalili et al., 2016). این آزمون توسط کومار (Kumar et al., 2009) با بررسی اثر خودهمبستگی در سری زمانی توسعه یافت و بعداً در مطالعات مختلفی مورد بررسی قرار گرفت (Kumar et al., 2009; Khaliq et al., 2009; Zamani et al., 2018; Tahroudi et al., 2019; Khozaymehnejad et al., 2019; Ramezani et al., 2020; Ahmadi et al., 2018&2022).

مدل جنگل تصادفی

مدل جنگل تصادفی به‌عنوان ابزار ماشین یادگیری بین محققین رونق یافته است. این مدل، یک مدل گروه پایه است که چند الگوریتم درختی را جهت پیش‌بینی مکرر از هر پدیده ترکیب می‌کند. یکی از ویژگی‌های این مدل، توانایی آن در یادگیری مدل‌های پیچیده و ارتباط غیرخطی بین متغیرهای توصیفی و متغیرهای وابسته است. بریمان (Breiman, 2001) با استفاده از قانون اعداد بزرگ نشان داد که هر چه تعداد درخت‌های یک جنگل افزایش می‌یابد، خطای پیش‌بینی جنگل، همگرا به یک مقدار می‌شود (Wang et al., 2005). 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 h_n ، 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):

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

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

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 y_n خروجی درخت n ام می‌باشد. برای به دست آوردن خروجی نهایی، متوسط همه پیش‌بینی‌های درخت‌ها محاسبه می‌شود (Breiman, 2001).

رگرسیون فرآیند گاوسی (GPR)

در سال‌های اخیر، محققان متعددی از مدل‌های شبیه‌سازی مبتنی بر یادگیری ماشین برای حل مسائل در زمینه‌های مختلف استفاده کرده‌اند که می‌توان به مطالعات (Iannace, Iannace et al (2018)

در مطالعات مختلف بررسی و تأیید شده است (Khalili et al., 2016; Ahmadi et al., 2022). نتایج بررسی روند تغییرات مقادیر مورد بررسی به همراه شیب تغییرات روند آن به صورت جدول ۲ ارائه گردید.

با توجه به نتایج ارائه شده در جدول ۲ می‌توان مشاهده کرد که تغییرات دبی جریان در مقیاس سالانه در تمامی ایستگاه‌های مورد بررسی کاهش می‌باشد. در مقیاس سالانه تغییرات کاهش دبی جریان در دو ایستگاه دیزج و چهریق علیا در سطح ۵ درصد معنی‌دار می‌باشد. روند تغییرات کاهش دبی جریان در ایستگاه تپیک در دوره آماری ۹۹-۱۳۵۰ معنی‌دار نمی‌باشد. با توجه به تغییرات شیب روند در داده‌های دبی جریان در ایستگاه‌های دیزج، تپیک و چهریق علیا، به ترتیب حدود ۲/۹۰، ۴/۱۵ و ۲/۴۶ مترمکعب بر ثانیه از متوسط دبی سالانه این ایستگاه‌ها کم شده است. این تغییرات در طی ۴۷ سال در ایستگاه‌های مورد مطالعه رخ داده است. تغییرات دبی جریان در ایستگاه‌های مورد بررسی سبب افزایش تغییرات پارامترهای کیفی در منطقه مورد مطالعه بوده است. تغییرات افزایشی در مقادیر کیفی مورد بررسی (EC، TDS و بی‌کربنات‌ها) در هر سه ایستگاه مورد بررسی مطابق با جدول ۲ قابل مشاهده می‌باشد.

نتایج شبیه‌سازی‌ها به شرح زیر استفاده شده است:

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (O_i - P_i)^2}{N}} \quad (7)$$

$$NSE = 1 - \frac{\sum_{i=1}^N (O_i - P_i)^2}{\sum_{i=1}^N (O_i - \bar{O}_i)^2} \quad (8)$$

که در آن N برابر با تعداد داده‌ها، O_i و P_i به ترتیب مقادیر مشاهداتی و شبیه‌سازی شده بوده و \bar{O}_i نیز متوسط داده‌های مشاهداتی می‌باشد.

نتایج و بحث

نتایج بررسی تغییرات روند مقادیر کمی و کیفی بررسی

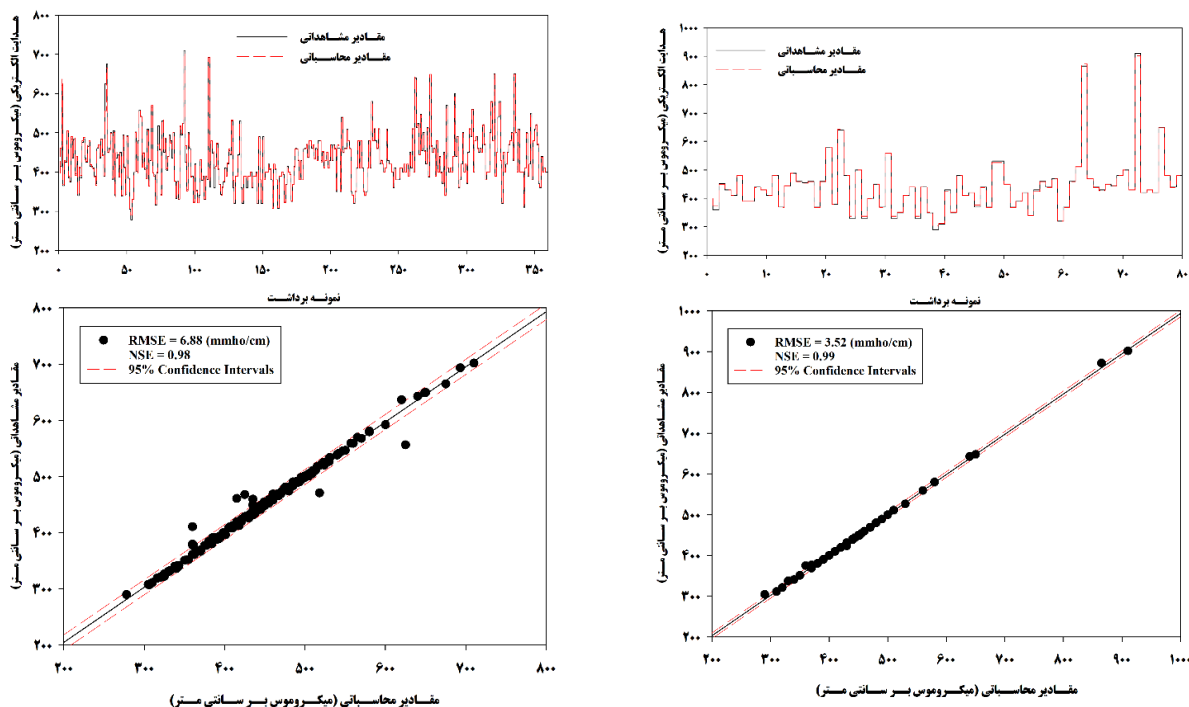
روند تغییرات مقادیر کیفی (بی‌کربنات‌ها، هدایت الکتریکی و کل مواد جامد محلول) و همچنین مقادیر دبی جریان متناظر (مترمکعب بر ثانیه) در ایستگاه‌های مورد بررسی به همراه شیب تغییرات روند به ترتیب با استفاده از آزمون‌های من-کندال اصلاح شده و شیب سن بررسی شد. حذف اثرات خودهمبستگی در داده‌ها قبل از بررسی روند، سبب می‌شود نتایج بررسی روند به صورت واقعی بیان شود. این روش

جدول ۲- روند تغییرات مقادیر مورد بررسی و شیب تغییرات آن در دوره آماری ۹۹-۱۳۵۰

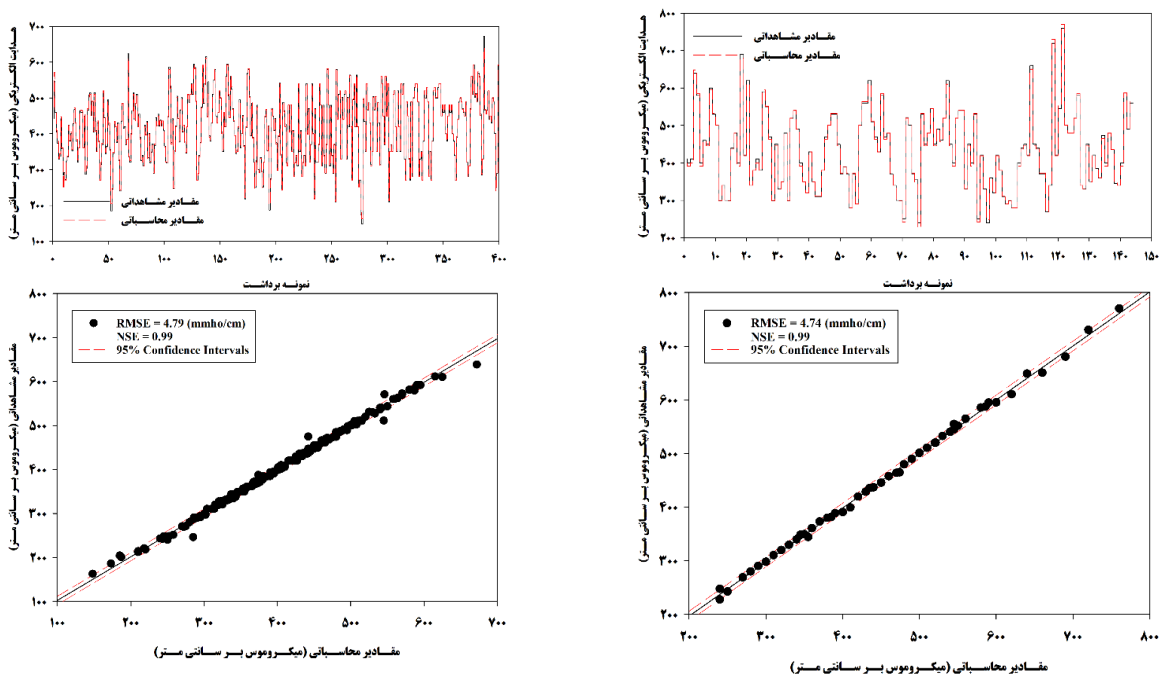
Station	Chehriq Olya					Tapik					Dizaj				
	(l/su) SCL	EC (mmho/cm)	HCO ₃ ⁻ (mg/l)	دبی (m ³ /s)	TDS (mg/l)	EC (mmho/cm)	HCO ₃ ⁻ (mg/l)	دبی (m ³ /s)	TDS (mg/l)	EC (mmho/cm)	HCO ₃ ⁻ (mg/l)	دبی (m ³ /s)			
Slope	-۰/۱۰	-۰/۲۸	-۰/۰۰	-۰/۰۵	-۰/۶۸	-۰/۵۱	-۰/۰۷	-۰/۰۹	-۰/۵۷	-۰/۸۵	-۰/۰۳	-۰/۰۶			
MK	-۰/۳۶	-۰/۹۲	-۰/۸۲	-۲/۳۵	۱/۵۲	-۰/۹۰	۱/۲۰	-۱/۰۴	۱/۷۹	۱/۷۲	-۰/۰۳	-۰/۰۲			

و تغییر در یکی سبب تغییر در دیگری می‌شود. در این مطالعه جهت شبیه‌سازی مقادیر EC و TDS در هر ایستگاه از داده‌های بی‌کربنات-ها و دبی جریان متناظر در دوره آماری ۹۹-۱۳۵۰ استفاده شد. داده‌های مورد بررسی در دو مجموعه آموزش و آزمایش به مدل‌های مورد بررسی معرفی شدند. در این مطالعه از دو مدل جنگل تصادفی و رگرسیون فرایند گوسی استفاده شد. جهت شبیه‌سازی‌ها، از مجموع داده‌های مورد بررسی، در ایستگاه‌های چهریق علیا، دیزج و تپیک به ترتیب ۴۰۰، ۳۶۰ و ۴۰۰ داده برای مجموعه آموزش و تعداد ۸۰، ۱۴۴ و ۱۱۲ داده برای مجموعه آزمایش در نظر گرفته شد.

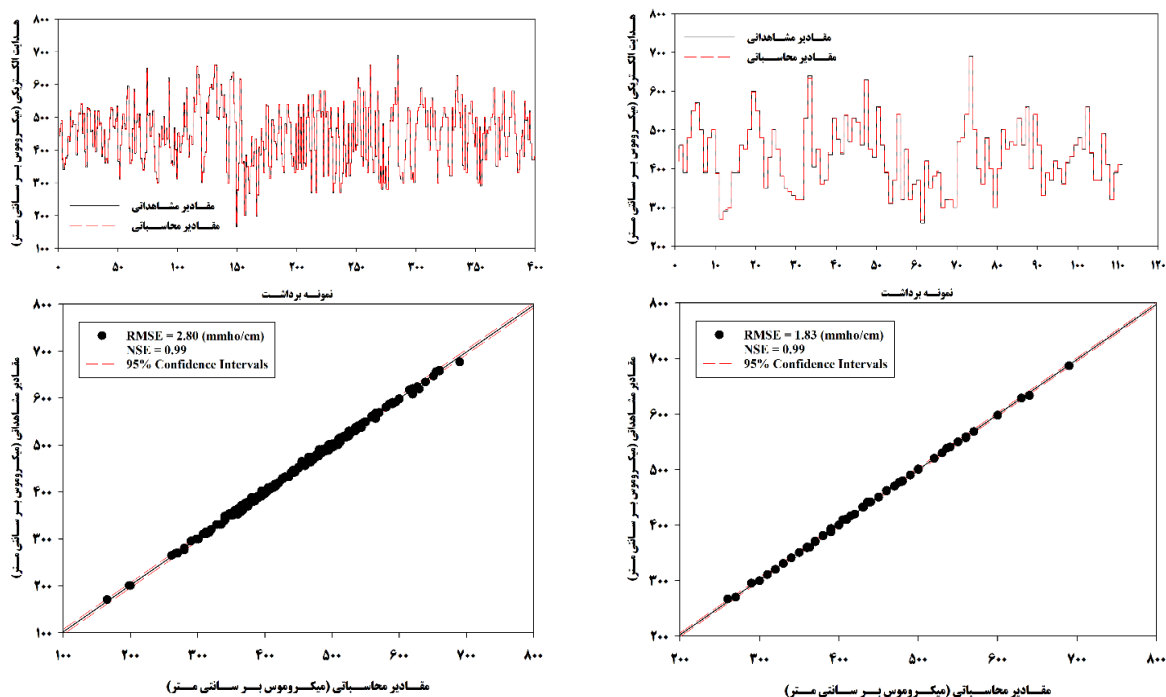
تغییرات افزایشی در هدایت الکتریکی، کل مواد جامد محلول و بی‌کربنات‌های موجود در جریان رودخانه معنی‌دار نمی‌باشد. میزان تغییرات ۴۷ ساله در مقادیر هدایت الکتریکی و کل مواد جامد محلول در ایستگاه دیزج با توجه به آماره شیب سن (شیب خط روند)، به ترتیب افزایش ۴۰ میکروموس بر سانتی‌متر و ۲۷ میلی‌گرم بر لیتر می‌باشد. این تغییرات در ایستگاه تپیک برابر با ۲۴ میکروموس بر سانتی‌متر و ۳۲ میلی‌گرم بر لیتر و در ایستگاه چهریق علیا برابر با ۱۳ میکروموس بر سانتی‌متر و ۴ میلی‌گرم بر لیتر است. تغییرات بی-کربنات‌ها در طی دوره آماری مورد بررسی نیز ناچیز می‌باشد. نتایج بررسی روند تغییرات مقادیر کمی و کیفی در سطح منطقه مورد مطالعه نشان داد که بین مقادیر کیفی و کمی رابطه خطی برقرار است



شکل ۲- نتایج شبیه‌سازی مقادیر هدایت الکتریکی در ایستگاه چهریق علیا با استفاده از مدل جنگل تصادفی در دوره آماری ۹۹-۱۳۵۰ (چپ: فاز آموزش، راست: فاز آزمایش)



شکل ۳- نتایج شبیه‌سازی مقادیر هدایت الکتریکی در ایستگاه دیزج با استفاده از مدل جنگل تصادفی در دوره آماری ۹۹-۱۳۵۰ (چپ: فاز آموزش، راست: فاز آزمایش)



شکل ۴- نتایج شبیه‌سازی مقادیر هدایت الکتریکی در ایستگاه تپیک با استفاده از مدل جنگل تصادفی در دوره آماری ۹۹-۱۳۵۰ (چپ: فاز آموزش، راست: فاز آزمایش)

میکروموس بر سانتی‌متر می‌باشد. میزان کارایی مدل بر اساس آماره نش-ساتکلیف (NSE) در هر دو فاز آموزش و آزمایش در ایستگاه چهریق علیا ۹۸ درصد و بیشتر می‌باشد. در این ایستگاه مطابق با شکل ۲، چندین مورد در مرحله آموزش از حدود اطمینان ۹۵ درصد شبیه‌سازی فراتر رفته است که این موارد در مرحله آزمایش کمتر می‌باشد. این الگوی تغییرات در ایستگاه‌های دیگر نیز مشاهده می‌شود. با توجه به شکل ۳ که تغییرات هدایت الکتریکی مقادیر شبیه‌سازی شده توسط مدل RF و مقادیر مشاهداتی را در ایستگاه دیزج نشان می‌دهد، می‌توان تطابق خوبی بین مقادیر مشاهداتی و شبیه‌سازی شده مشاهده کرد. مدل RF در مرحله آموزشی متوسط مقادیر شبیه‌سازی شده را ۰/۰۱ درصد بیشتر از متوسط مقادیر مشاهداتی و متوسط مقادیر شبیه‌سازی شده در مرحله آزمایش را ۰/۳۸ درصد کمتر از متوسط مقادیر مشاهداتی متناظر شبیه‌سازی کرده است. در مرحله آموزش و آزمایش، میزان خطای شبیه‌سازی مقادیر هدایت الکتریکی در ایستگاه دیزج به ترتیب ۴/۸ و ۴/۷ میکروموس بر سانتی‌متر می‌باشد که با توجه به دامنه تغییرات داده‌های مورد بررسی، میزان خطای قابل قبولی می‌باشد. کارایی مدل RF در شبیه‌سازی مقادیر هدایت الکتریکی نیز در فاز آموزش و آزمایش بیش از ۹۹ درصد است که توانایی بالای مدل RF در شبیه‌سازی سه متغیره مقادیر کمی و کیفی

نتایج شبیه‌سازی هدایت الکتریکی ایستگاه‌های مورد بررسی با استفاده از مدل جنگل تصادفی

با استفاده از مجموعه داده‌های آموزش و آزمایش برای ایستگاه‌های مورد مطالعه، شبیه‌سازی سه متغیره هدایت الکتریکی مورد مطالعه قرار گرفت. مدل جنگل تصادفی برای مجموعه داده‌های آموزشی پیاده‌سازی شده و در نهایت، دقت مدل در شبیه‌سازی هدایت الکتریکی مورد بررسی قرار گرفت. نتایج شبیه‌سازی مقادیر هدایت الکتریکی در ایستگاه‌های مورد مطالعه در دو فاز آموزش و آزمایش به شرح شکل‌های ۲ تا ۴ ارائه گردید.

با توجه به شکل‌های ۲ تا ۴ می‌توان تطابق خوبی بین مقادیر مشاهداتی و شبیه‌سازی شده توسط مدل جنگل تصادفی در ایستگاه‌های مختلف مشاهده کرد. در ایستگاه چهریق علیا، متوسط مقادیر شبیه‌سازی شده توسط مدل RF در فاز آموزش ۰/۱۰ درصد و در فاز آزمایش ۰/۰۷ درصد بیشتر از مقادیر مشاهداتی می‌باشد که نشان دهنده دقت بالای مدل RF در شبیه‌سازی مقادیر هدایت الکتریکی در این ایستگاه می‌باشد. در ایستگاه چهریق علیا در مرحله آموزش مطابق با شکل ۴-۱، مقدار خطای شبیه‌سازی هدایت الکتریکی با استفاده از مدل RF بر اساس معیار RMSE، برابر با ۶/۸۸ میکروموس بر سانتی‌متر و در مرحله آزمایش نیز برابر با ۳/۵۲

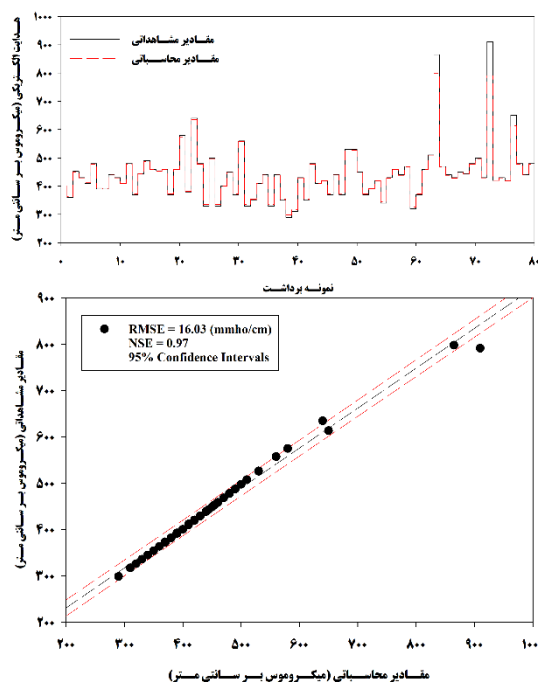
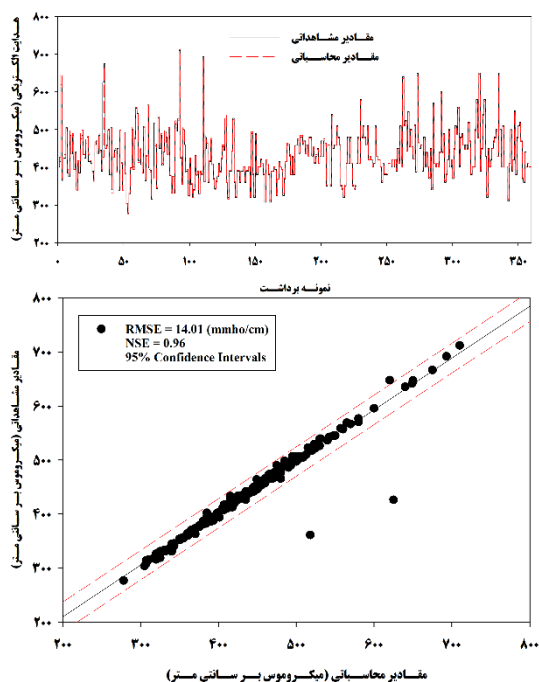
مدل در شبیه‌سازی هدایت الکتریکی مورد بررسی قرار گرفت. نتایج شبیه‌سازی مقادیر هدایت الکتریکی منتج از مدل رگرسیون فرآیند گاوسی در ایستگاه‌های مورد مطالعه در دو فاز آموزش و آزمایش به شرح شکل‌های ۵ تا ۷ ارائه گردید.

همانند مدل جنگل تصادفی، با توجه به شکل‌های ۵ تا ۷ می‌توان برآزش مناسی بین مقادیر مشاهداتی و شبیه‌سازی شده توسط مدل رگرسیون فرآیند گاوسی در ایستگاه‌های مورد مطالعه مشاهده کرد. در ایستگاه چهریق علیا مطابق با شکل ۵ در فاز آموزش چند نقطه خارج از محدوده اطمینان ۹۵ درصد قرار دارد که کم تخمینی را نشان می‌دهد. متوسط مقادیر شبیه‌سازی شده توسط مدل GPR در فاز آموزش مشابه با متوسط مقادیر مشاهداتی بوده و اختلاف کمی دارد. در فاز آزمایش ۶۰/۰ درصد متوسط مقادیر شبیه‌سازی شده کمتر از متوسط مقادیر مشاهداتی می‌باشد که نشان دهنده دقت و عملکرد بالای مدل GPR در شبیه‌سازی مقادیر هدایت الکتریکی در این ایستگاه می‌باشد. مطابق با شکل ۵، در ایستگاه چهریق علیا در مرحله آموزش مقدار خطای شبیه‌سازی هدایت الکتریکی با استفاده از مدل GPR برابر با ۱۴/۰۱ میکروموس بر سانتی‌متر و در مرحله آزمایش نیز برابر با ۱۶/۰۲ میکروموس بر سانتی‌متر می‌باشد.

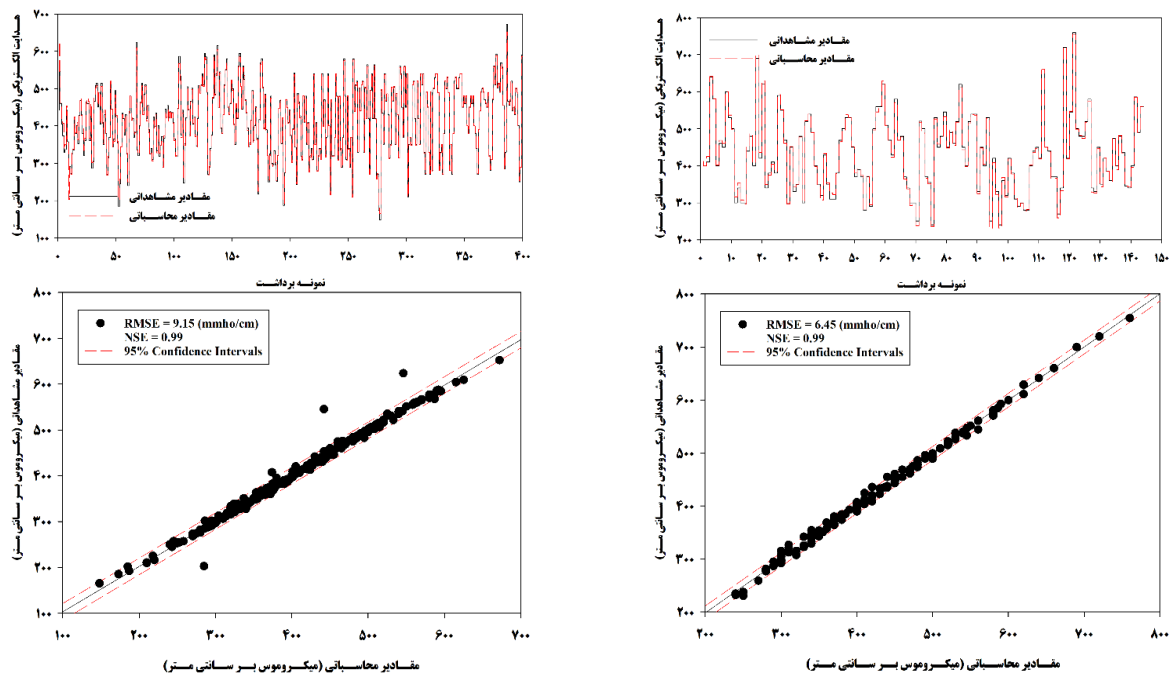
را نشان می‌دهد. از آنجایی که در شبیه‌سازی مقادیر مورد بررسی از متغیر دبی جریان و بی‌کربنات‌ها نیز استفاده شد، دقت بالا در شبیه‌سازی‌ها حاصل شد. زیرا هم متغیر دبی جریان و هم مقدار بی‌کربنات‌های متناظر با مقادیر EC همبستگی خوبی با داده‌های EC دارند. شکل ۴ نیز مقادیر مشاهداتی و شبیه‌سازی شده هدایت الکتریکی را در ایستگاه تپیک با توجه به تغییرات دبی جریان و مقادیر بی‌کربنات‌ها نشان می‌دهد. در این ایستگاه میزان خطا در فاز آموزش و آزمایش به ترتیب ۲/۸ و ۱/۸ میکروموس بر سانتی‌متر با توجه به مدل RF و ورودی‌های آن برآورد گردید که نسبت به ایستگاه‌های دیگر، میزان خطای کمتری در برآورد هدایت الکتریکی کسب کرده است. مدل RF در این ایستگاه متوسط مقادیر شبیه‌سازی شده در فاز آموزش را ۰/۰۶ درصد کمتر و متوسط مقادیر شبیه‌سازی شده در فاز آزمایش را ۰/۱۱ درصد بیشتر برآورد کرده است.

نتایج شبیه‌سازی هدایت الکتریکی ایستگاه‌های مورد بررسی با استفاده از مدل رگرسیون فرآیند گاوسی

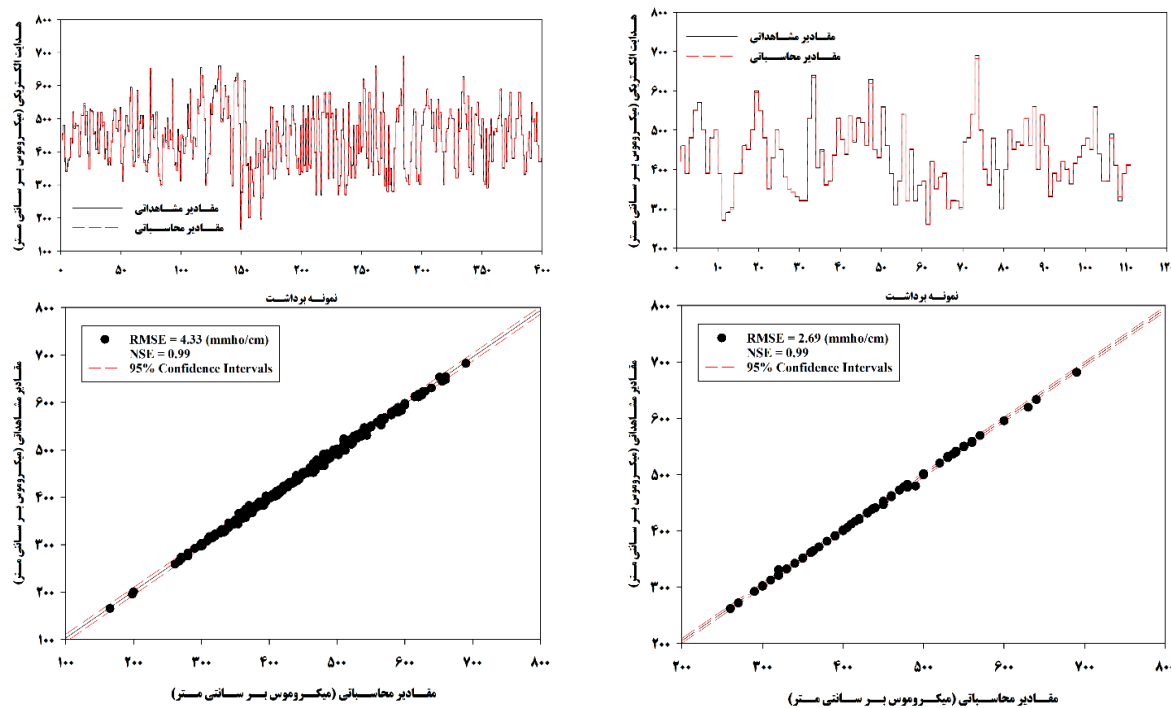
در این بخش مدل رگرسیون فرآیند گاوسی برای مجموعه داده‌های آموزشی پیاده‌سازی شده و همانند مدل جنگل تصادفی، دقت



شکل ۵- نتایج شبیه‌سازی مقادیر هدایت الکتریکی در ایستگاه چهریق علیا با استفاده از مدل رگرسیون فرآیند گاوسی در دوره آماری ۹۹-۱۳۵۰ (چپ: فاز آموزش، راست: فاز آزمایش)



شکل ۶- نتایج شبیه‌سازی مقادیر هدایت الکتریکی در ایستگاه دیزج با استفاده از مدل رگرسیون فرآیند گاوسی در دوره آماری ۹۹-۱۳۵۰ (چپ: فاز آموزش، راست: فاز آزمایش)



شکل ۷- نتایج شبیه‌سازی مقادیر هدایت الکتریکی در ایستگاه تپیک با استفاده از مدل رگرسیون فرآیند گاوسی در دوره آماری ۹۹-۱۳۵۰ (چپ: فاز آموزش، راست: فاز آزمایش)

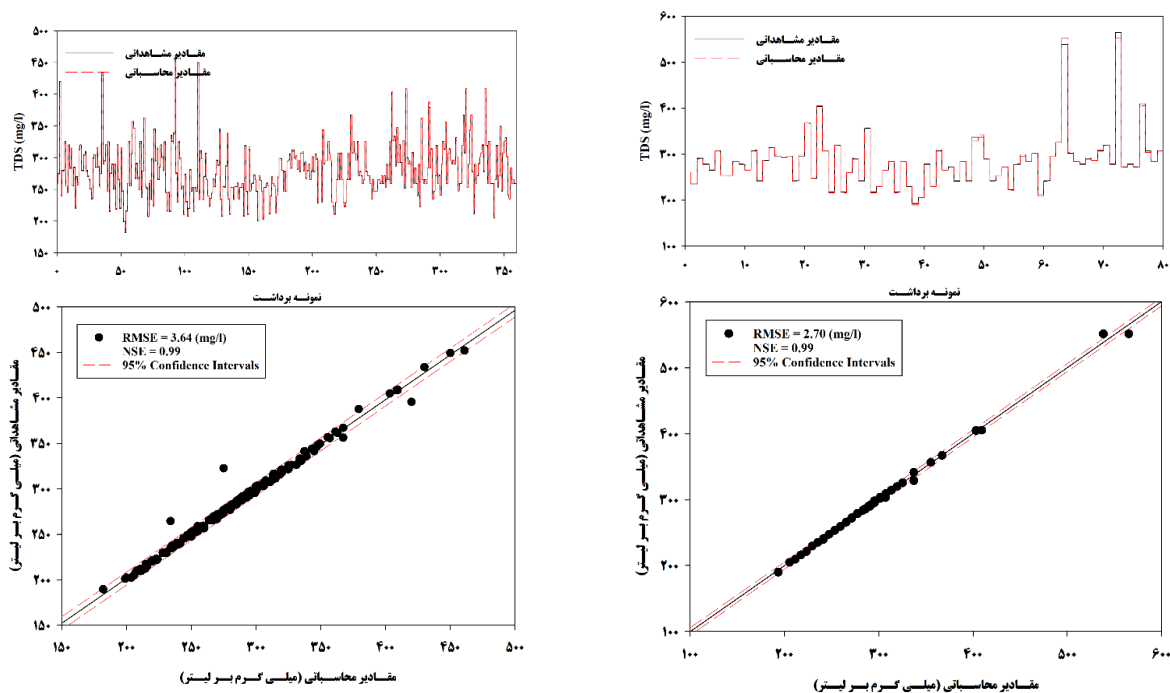
همانند دو ایستگاه چهریق علیا و دیزج، نتایج نشان دهنده برتری مدل RF نسبت به مدل GPR در شبیه‌سازی سه متغیره مقادیر هدایت الکتریکی می‌باشد.

نتایج شبیه‌سازی کل مواد جامد محلول ایستگاه‌های مورد بررسی با استفاده از مدل جنگل تصادفی

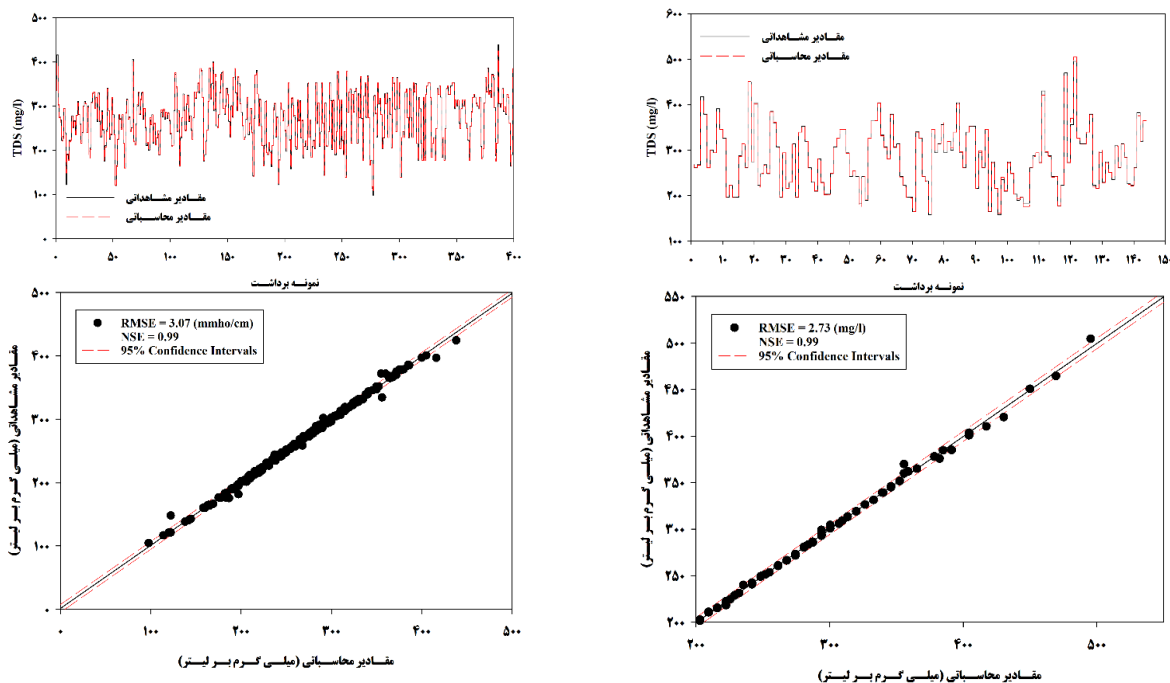
با استفاده از مقادیر دبی جریان، هدایت الکتریکی و بی‌کربنات-های اندازه‌گیری شده در جریان رودخانه در محل ایستگاه‌های مورد بررسی، شبیه‌سازی مقادیر TDS در دو فاز آموزش و آزمایش با استفاده از مدل جنگل تصادفی بررسی و نتایج شبیه‌سازی مقادیر TDS در ایستگاه‌های مورد بررسی به صورت شکل‌های ۸ تا ۱۰ ارائه گردید. با توجه به شکل‌های ۸ تا ۱۰ می‌توان تطابق خوبی بین مقادیر مشاهداتی و شبیه‌سازی شده TDS در تمامی ایستگاه‌ها مشاهده کرد. با توجه به شکل ۸ که نتایج شبیه‌سازی مقادیر TDS در ایستگاه چهریق علیا را نشان می‌دهد می‌توان مشاهده کرد که در مرحله آموزش سه مورد از نمونه‌های شبیه‌سازی شده خارج از محدوده اطمینان ۹۵ درصد شبیه‌سازی واقع شده‌اند که بیانگر دو مورد بیش تخمینی و یک مورد کم تخمینی است. در مرحله آموزش میزان خطای مدل ۳/۶۵ میلی‌گرم بر لیتر برآورد شد. در این فاز آموزش متوسط مقادیر شبیه‌سازی شده TDS توسط مدل جنگل تصادفی نسبت به مقادیر مشاهداتی ۰/۰۳ درصد افزایش داشته که بسیار ناچیز می‌باشد. این درصد در مرحله آزمایش نزدیک به صفر برآورد شد. میزان خطا در فاز آموزش نیز حدود ۲/۷۰ میلی‌گرم بر لیتر محاسبه شده و با توجه به شکل ۸، دو مورد از نمونه‌های شبیه‌سازی شده خارج از محدوده اطمینان ۹۵ درصد قرار گرفته‌اند. علاوه بر این نمودارهای ارائه شده بیانگر تطبیق خوب بین مقادیر مشاهداتی و شبیه‌سازی شده در این ایستگاه می‌باشد. با توجه به شکل ۹ می‌توان مشاهده کرد که در ایستگاه دیزج نسبت به ایستگاه چهریق علیا، در فاز آموزش تعداد نمونه‌های بیشتری خارج از محدوده شبیه‌سازی شده توسط مدل جنگل تصادفی قرار گرفته است. میزان خطای شبیه‌سازی در فاز آموزش بر اساس معیار RMSE در ایستگاه دیزج برابر با ۳/۰۷ میلی‌گرم بر لیتر می‌باشد. این معیار در فاز آموزش عدد ۲/۷۳ میلی‌گرم بر لیتر را نشان می‌دهد. مدل جنگل تصادفی به‌خوبی توانسته است مقادیر کمینه و بیشینه TDS را در ایستگاه دیزج شبیه‌سازی کند. بر اساس مدل جنگل تصادفی، متوسط مقادیر شبیه‌سازی شده TDS در فاز آموزش و آزمایش به ترتیب برابر با ۲۷۱/۲۱ و ۲۸۲/۷۰ تخمین زده شد که نسبت به مقادیر مشاهداتی به ترتیب افزایش ۰/۰۶ و کاهش ۰/۱۷ را نشان می‌دهد. نتایج شبیه‌سازی مقادیر TDS در ایستگاه تپیک با استفاده از مدل جنگل تصادفی به صورت شکل ۱۰ ارائه شده است که به‌خوبی تطبیق مقادیر مشاهداتی و شبیه‌سازی شده را در این ایستگاه نشان می‌دهد.

در مقایسه با مدل RF، میزان خطای مدل GPR در مرحله آموزش و آزمایش به ترتیب ۱۰۴ و ۳۵۶ درصد بیشتر می‌باشد. میزان کارایی مدل در فاز آموزش و آزمایش به ترتیب ۹۶ و ۹۷ درصد در ایستگاه چهریق علیا می‌باشد که نسبت به مدل RF حدود ۳ درصد کمتر است. نتایج شبیه‌سازی هدایت الکتریکی در ایستگاه چهریق علیا نشان داد که مدل RF نتایج بهتری نسبت به مدل GPR ارائه کرده است که نشان دهنده دقت و کارایی بالای مدل RF نسبت به مدل GPR می‌باشد.

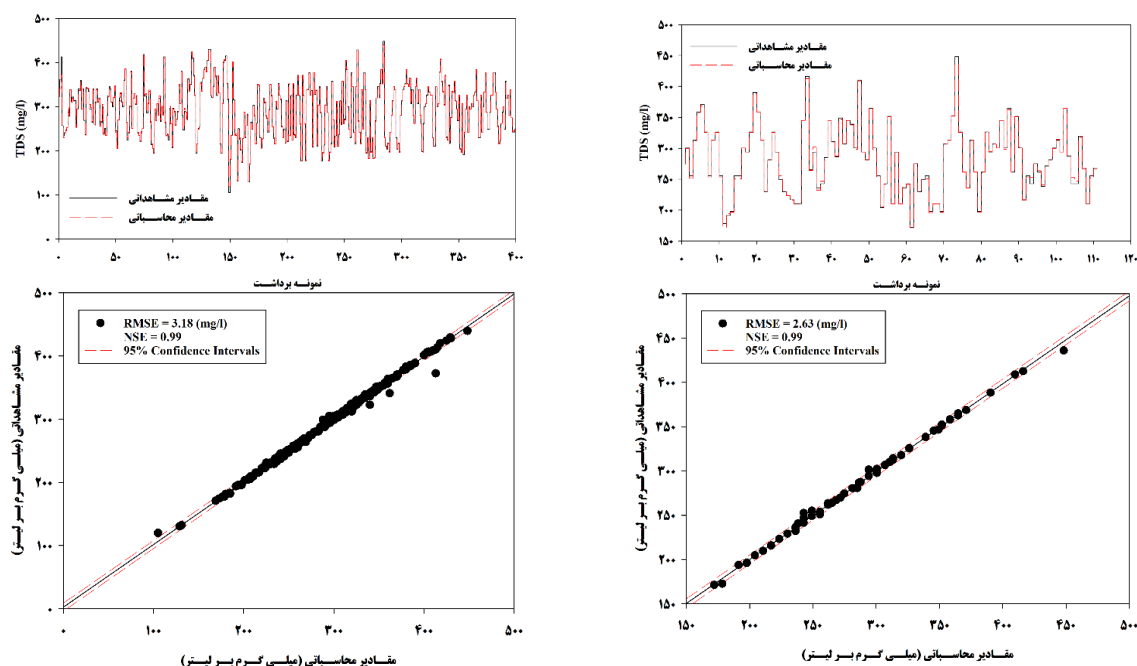
با توجه به شکل ۶ که تغییرات هدایت الکتریکی مقادیر شبیه‌سازی شده توسط مدل GPR و مقادیر مشاهداتی را در ایستگاه دیزج نشان می‌دهد، می‌توان مشاهده کرد که کمی بیش تخمین در شبیه‌سازی مقادیر هدایت الکتریکی وجود دارد که از محدوده ۹۵ درصد اطمینان در فاز آموزش فراتر است. اما به‌طور کلی در دو فاز آموزش و آزمایش در ایستگاه دیزج تطابق خوبی بین مقادیر مشاهداتی و شبیه‌سازی شده مشاهده می‌شود. در ایستگاه دیزج، متوسط مقادیر شبیه‌سازی شده توسط مدل GPR در فاز آموزش و آزمایش به ترتیب برابر با ۴۱۵/۰۸ و ۴۳۴/۵۶ میکروموس بر سانتی‌متر می‌باشد که نسبت به حالت مشاهداتی در فاز آموزش تفاوت چندانی نداشته و در فاز آموزش حدود ۰/۲۴ درصد بیشتر از متوسط مقادیر مشاهداتی می‌باشد. در ایستگاه دیزج، در مرحله آموزش و آزمایش، میزان خطای شبیه‌سازی مقادیر هدایت الکتریکی با استفاده از مدل GPR به ترتیب برابر با ۹/۱۵ و ۶/۴۶ میکروموس بر سانتی‌متر می‌باشد که در مقایسه با نتایج ارائه شده توسط مدل RF در فاز آموزش و آزمایش به ترتیب ۹۱ و ۳۶ درصد بیشتر از مدل RF می‌باشد. کارایی مدل GPR و مدل RF در هر دو فاز آموزش و آزمایش تفاوت چندانی نداشته و عملکرد هر دو مدل در شبیه‌سازی مقادیر هدایت الکتریکی قابل قبول می‌باشد. همانند ایستگاه چهریق علیا، دقت و عملکرد مدل RF در شبیه‌سازی مقادیر هدایت الکتریکی در دو فاز آموزش و آزمایش بهتر از مدل GPR می‌باشد. شکل ۷ نیز مقادیر مشاهداتی و شبیه‌سازی شده هدایت الکتریکی را در ایستگاه را با استفاده از مدل GPR و به شرط وقوع مقادیر بی‌کربنات‌ها و دبی جریان نشان می‌دهد. برخلاف دو ایستگاه چهریق علیا و دیزج، در ایستگاه تپیک تمامی مقادیر شبیه‌سازی شده هدایت الکتریکی در محدوده اطمینان ۹۵ درصد قرار دارد. متوسط مقادیر شبیه‌سازی شده در مقایسه با متوسط مقادیر مشاهداتی در ایستگاه تپیک با توجه به مدل GPR در فاز آموزش و آزمایش به ترتیب ۰/۲۸ درصد کمتر و ۰/۱۹ درصد بیشتر می‌باشد. نتایج شبیه‌سازی مقادیر هدایت الکتریکی در ایستگاه تپیک نشان داد که در این ایستگاه میزان خطا در فاز آموزش و آزمایش به ترتیب ۴/۳۳ و ۲/۶۹ میکروموس بر سانتی‌متر می‌باشد که نسبت به مدل RF ۵۵ درصد در فاز آموزش و ۴۷ درصد در فاز آموزش میزان خطای بیشتری دارد که



شکل ۸- نتایج شبیه‌سازی مقادیر TDS در ایستگاه چهریق علیا با استفاده از مدل جنگل تصادفی در دوره آماری ۹۹-۱۳۵۰ (چپ: فاز آموزش، راست: فاز آزمایش)



شکل ۹- نتایج شبیه‌سازی مقادیر TDS در ایستگاه دیزج با استفاده از مدل جنگل تصادفی در دوره آماری ۹۹-۱۳۵۰ (چپ: فاز آموزش، راست: فاز آزمایش)



شکل ۱۰- نتایج شبیه‌سازی مقادیر TDS در ایستگاه تپیک با استفاده از مدل جنگل تصادفی در دوره آماری ۹۹-۱۳۵۰ (چپ: فاز آموزش، راست: فاز آزمایش)

ارائه شد. شکل ۱۱ نتایج شبیه‌سازی مقادیر TDS در ایستگاه چهریق علیا را با استفاده از مدل رگرسیون فرآیند گاوسی در دو فاز آموزش و آزمایش نشان می‌دهد. با توجه به این که چندین مورد از مقادیر شبیه‌سازی شده خارج از محدوده اطمینان ۹۴ درصد در هر دو فاز آموزش و آزمایش قرار دارند، با این حال بین مقادیر مشاهده‌ای و شبیه‌سازی شده تطابق مناسبی مشاهده می‌شود. در فاز آموزش و آزمایش به ترتیب میزان خطای ۸/۰۷ و ۹/۹۳ میلی‌گرم بر لیتر با توجه به آماره RMSE برآورد شد و کارایی بیش از ۹۶ درصد مدل رگرسیون فرآیند گاوسی در شبیه‌سازی مقادیر TDS در ایستگاه چهریق علیا تأیید شد. در این ایستگاه متوسط مقادیر TDS مشاهده‌ای در فاز آموزش و آزمایش به ترتیب برابر با ۲۷۸/۸ و ۲۸۴/۶ میلی‌گرم بر لیتر می‌باشد که مدل رگرسیون فرآیند گاوسی متوسط مقادیر شبیه‌سازی شده در فاز آموزش و آزمایش را به ترتیب ۲۷۸/۹ و ۲۸۵/۱ برآورد کرده است که افزایش ۰/۱۸ درصدی را در فاز آموزش نشان می‌دهد. مقایسه نتایج مدل RF و GPR در شبیه‌سازی مقادیر TDS در ایستگاه چهریق علیا نشان داد که مدل RF در مرحله آموزش توانسته است ۱۲۱ درصد نتایج میزان خطا (RMSE) را نسبت به مدل GPR بهبود بخشد (کاهش دهد). آماره RMSE در شبیه‌سازی مقادیر TDS در ایستگاه چهریق علیا با استفاده از مدل RF، عدد ۳/۶۴ میلی‌گرم بر لیتر را نشان داد. این عدد در فاز آزمایش ۲/۶۹

هم در فاز آموزش و هم در فاز آزمایش تعدادی از نمونه‌های شبیه‌سازی ده خارج از محدوده اطمینان ۹۵ درصد واقع شده‌اند، با این حال کارایی مدل در هر دو فاز آموزش و آزمایش ۹۹ درصد و قابل قبول می‌باشد. میزان خطای شبیه‌سازی مقادیر TDS در ایستگاه تپیک در دو فاز آموزش و آزمایش به ترتیب برابر با ۳/۱۸ و ۲/۶۳ میلی‌گرم بر لیتر می‌باشد. مدل جنگل تصادفی در فاز آموزش متوسط مقادیر شبیه‌سازی شده را مشابه با متوسط مقادیر مشاهده‌ای برآورد کرد. در فاز آزمایش، متوسط مقادیر شبیه‌سازی شده توسط مدل جنگل تصادفی حدود ۰/۱۷ درصد کمتر از متوسط مقادیر مشاهده‌ای برآورد شد.

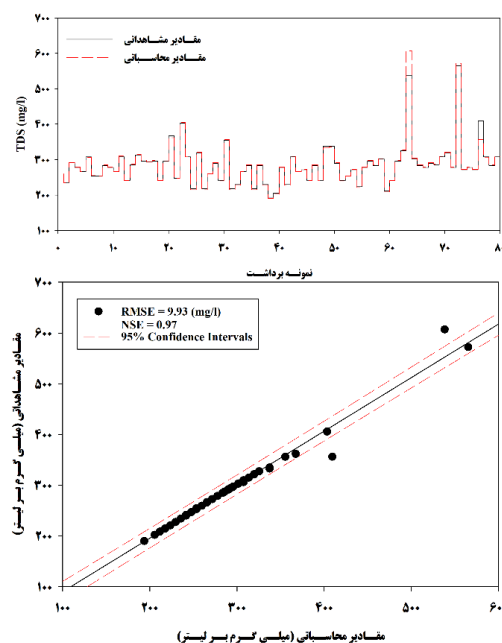
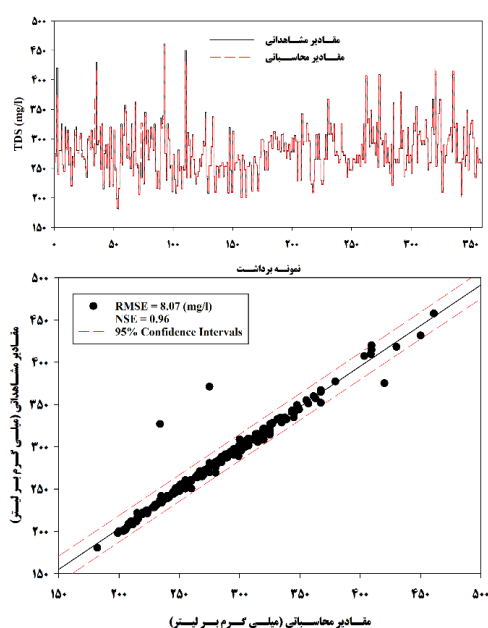
نتایج شبیه‌سازی کل مواد جامد محلول ایستگاه‌های مورد بررسی با استفاده از مدل رگرسیون فرآیند گاوسی

در این بخش جهت شبیه‌سازی مقادیر TDS در ایستگاه‌های مورد بررسی با توجه به مقادیر هدایت الکتریکی، بی‌کربنات‌ها و دی‌کربنات‌ها، از مدل رگرسیون فرآیند گاوسی استفاده شد. شبیه‌سازی مقادیر TDS با استفاده از مدل رگرسیون فرآیند گاوسی در دو فاز آموزش و آزمایش همانند مدل‌های دیگر انجام شد. نتایج شبیه‌سازی مقادیر TDS با در نظر گرفتن سایر پارامترهای ورودی (هدایت الکتریکی، بی‌کربنات‌ها و دی‌کربنات‌ها)، به صورت شکل‌های ۱۱ تا ۱۳

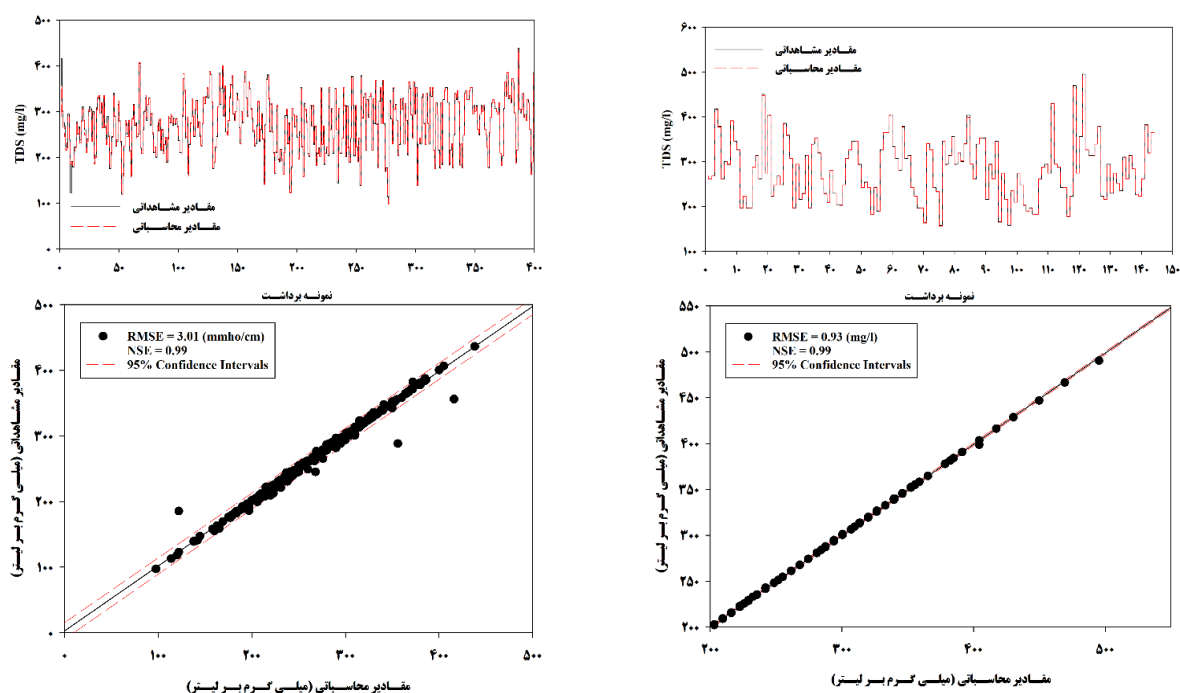
به‌عنوان معیار خطا در شبیه‌سازی مقادیر TDS در ایستگاه دیزج نشان می‌دهد. در مقایسه با مدل RF، بر خلاف ایستگاه چهریق علیا، در ایستگاه دیزج عملکرد و میزان خطای مدل GPR نسبت به مدل RF بهتر می‌باشد. نتایج مقایسه دو مدل RF و GPR در شبیه‌سازی مقادیر TDS در ایستگاه دیزج نشان داد که مدل GPR نسبت به مدل RF در فاز آموزش و آزمایش میزان خطای شبیه‌سازی را به ترتیب حدود ۲ و ۶۶ درصد بهبود داده است. میزان کارایی هر دو مدل در شبیه‌سازی مقادیر TDS در ایستگاه دیزج ۹۹ درصد و قابل قبول برآورد شد. مدل GPR نسبت به مدل RF در فاز آزمایش عملکرد بسیار خوبی نسبت به مدل RF ارائه کرده است. مطابق شکل ۱۳ نیز می‌توان نتایج شبیه‌سازی مقادیر TDS ایستگاه تپیک را با استفاده از مدل GPR در دو فاز آموزش و آزمایش مشاهده کرد. عملکرد مدل GPR در هر دو فاز آموزش و آزمایش بسیار رضایت‌بخش می‌باشد. تقریباً تمامی مقادیر شبیه‌سازی شده در محدوده اطمینان ۹۵ درصد قرار گرفته و در هر دو فاز آموزش و آزمایش عملکرد رضایت‌بخش ۹۹ درصد را ارائه کرده است.

میلی‌گرم بر لیتر می‌باشد. در فاز آزمایش نیز مدل RF توانسته است میزان خطا را به‌صورت چشمگیری حدود ۲۶۸ درصد بهبود بخشد. کارایی مدل RF نیز در مقایسه با مدل GPR در دو فاز آموزش و آزمایش ۹۹ درصد می‌باشد که کارایی بالای مدل RF را در شبیه‌سازی مقادیر TDS در ایستگاه چهریق علیا نشان می‌دهد.

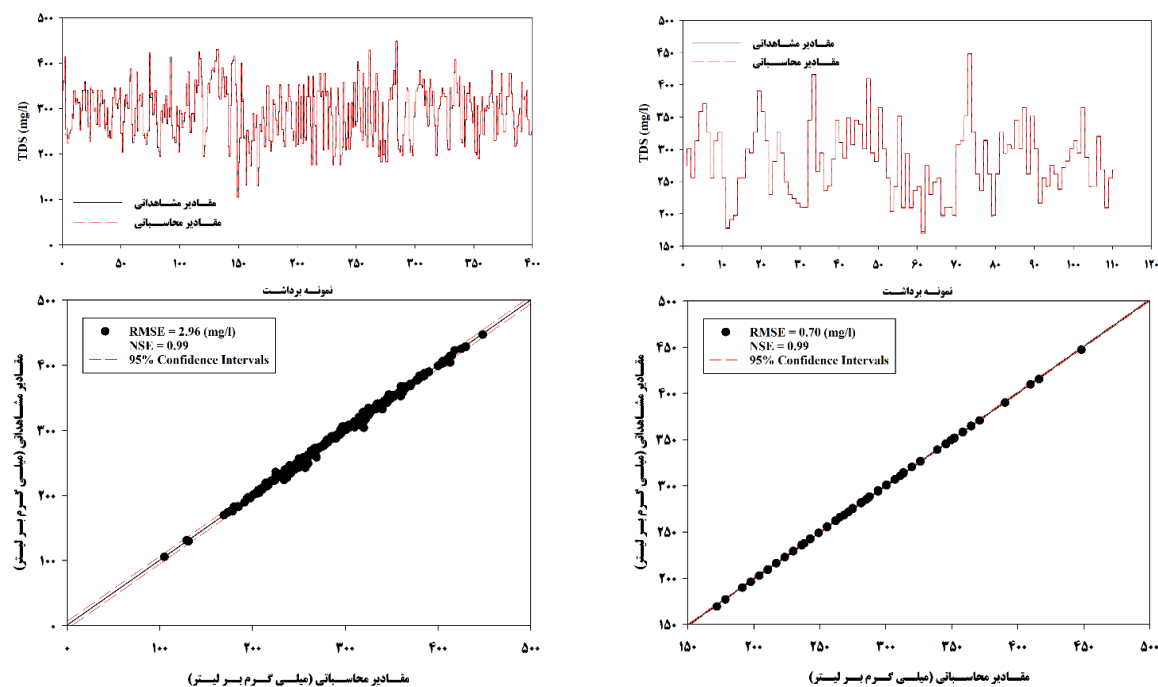
شکل ۱۲ نتایج شبیه‌سازی مقادیر TDS را در ایستگاه دیزج با توجه به مقادیر بی‌کربنات‌ها، هدایت الکتریکی و دبی جریان متناظر نشان می‌دهد. در فاز آموزش در این ایستگاه نسبت به ایستگاه چهریق علیا مقادیر خارج از محدوده اطمینان ۹۵ درصد کمتر می‌باشند. تطابق بسیار خوبی را می‌توان در فاز آزمایش در این ایستگاه مشاهده کرد. در شبیه‌سازی مقادیر TDS در ایستگاه دیزج در فاز آموزش و آزمایش متوسط مقادیر شبیه‌سازی شده به ترتیب حدود ۰/۰۹ و ۰/۱۲ درصد بیشتر و کمتر از متوسط مقادیر مشاهده‌ای می‌باشد. کارایی مدل در شبیه‌سازی مقادیر TDS در این ایستگاه ۹۹ درصد می‌باشد که کارایی مورد تأیید مدل GPR را در شبیه‌سازی مقادیر TDS نشان می‌دهد. شاخص RMSE نیز در دو فاز آموزش و آزمایش مطابق با شکل ۱۲ عدد ۳/۰۱ و ۰/۹۳ میلی‌گرم بر لیتر را



شکل ۱۱- نتایج شبیه‌سازی مقادیر TDS در ایستگاه چهریق علیا با استفاده از مدل رگرسیون فرآیند گاوسی در دوره آماری ۹۹-۱۳۵۰ (چپ: فاز آموزش، راست: فاز آزمایش)



شکل ۱۲- نتایج شبیه‌سازی مقادیر TDS در ایستگاه دیزج با استفاده از مدل رگرسیون فرآیند گاوسی در دوره آماری ۹۹-۱۳۵۰ (چپ: فاز آموزش، راست: فاز آزمایش)

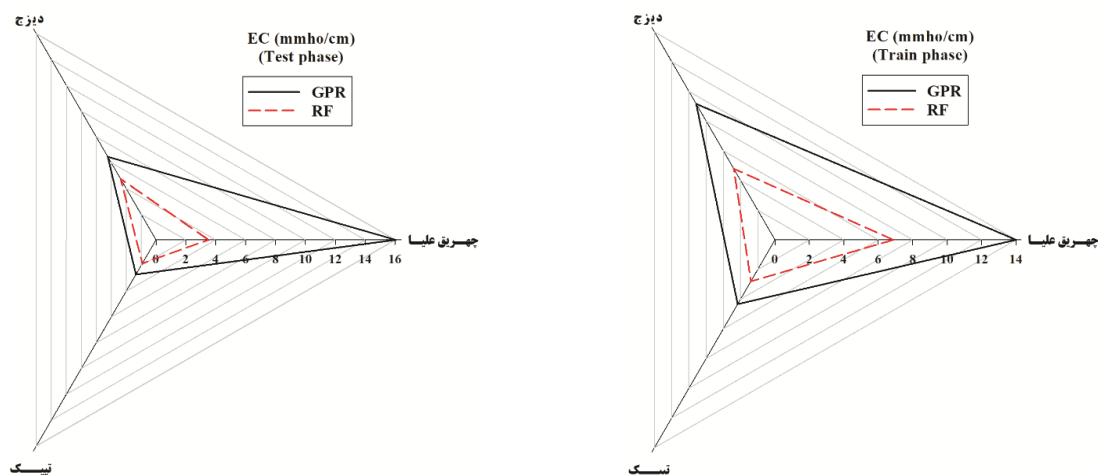


شکل ۱۳- نتایج شبیه‌سازی مقادیر TDS در ایستگاه تیبیک با استفاده از مدل رگرسیون فرآیند گاوسی در دوره آماری ۹۹-۱۳۵۰ (چپ: فاز آموزش، راست: فاز آزمایش)

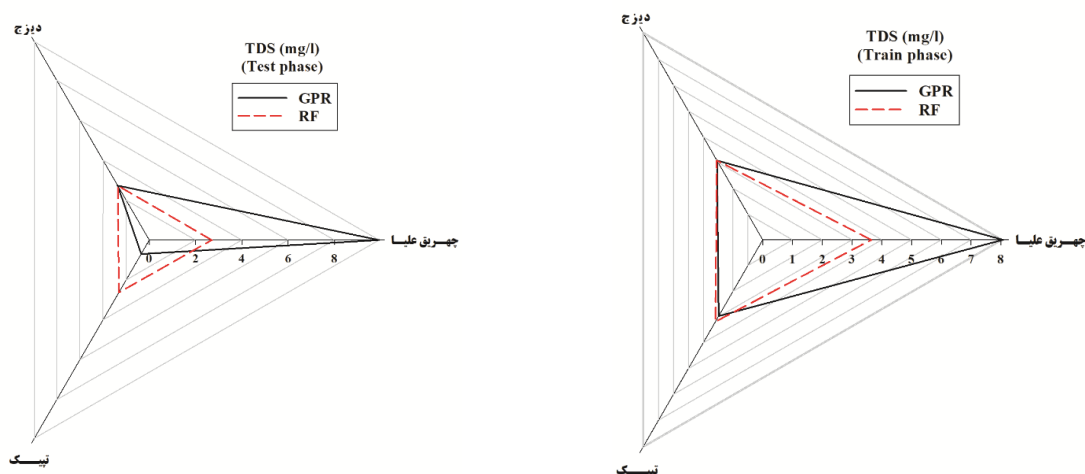
نسبت به مدل RF به ترتیب حدود ۷ و ۷۴ درصد بهبود داده است. همانند ایستگاه دیزج، مدل GPR توانست در فاز آزمایش میزان خطای شبیه‌سازی مقادیر TDS را به صورت چشمگیری افزایش داده است.

خلاصه نتایج برآورد خطا (RMSE) در شبیه‌سازی مقادیر هدایت الکتریکی و کل مواد جامد محلول با توجه به سایر پارامترهای ورودی (کل مواد جامد محلول، دبی جریان و بی‌کربنات‌ها برای شبیه‌سازی مقادیر هدایت الکتریکی و دبی جریان، هدایت الکتریکی و بی‌کربنات‌ها برای شبیه‌سازی مقادیر کل مواد جامد محلول) به صورت شکل‌های ۱۴ تا ۱۷ ارائه شد.

میزان خطای شبیه‌سازی مقادیر TDS در دو فاز آموزش و آزمایش در ایستگاه تپیک با توجه به مدل GPR به ترتیب حدود ۲/۹۶ و ۰/۷۰ میلی‌گرم بر لیتر می‌باشد که نسبت به دو ایستگاه چهریق علیا و دیزج کمتر می‌باشد. با توجه به شکل ۱۳ می‌توان تطابق خوبی بین مقادیر مشاهده‌ای و شبیه‌سازی شده مشاهده کرد که بیانگر دقت و عملکرد قابل قبول مدل GPR در شبیه‌سازی مقادیر TDS در ایستگاه تپیک می‌باشد. در مقایسه با مدل RF، عملکرد و میزان خطای مدل GPR در شبیه‌سازی مقادیر TDS در ایستگاه تپیک بهتر و رضایت‌بخش می‌باشد. با مقایسه نتایج شبیه‌سازی مقادیر TDS در ایستگاه تپیک در دو فاز آموزش و آزمایش در هر دو مدل RF و GPR، نتایج نشان داد که مدل GPR نتایج آماره RMSE را



شکل ۱۴- نتایج آماره RMSE در شبیه‌سازی مقادیر EC در ایستگاه‌های مورد مطالعه (چپ: فاز آموزش، راست: فاز آزمایش)



شکل ۱۵- نتایج آماره RMSE در شبیه‌سازی مقادیر TDS در ایستگاه‌های مورد مطالعه (چپ: فاز آموزش، راست: فاز آزمایش)

مبتنی بر جنگل تصادفی و رگرسیون فرآیند گاوسی برای پیش‌بینی جریان ماهانه نیز مشاهده کرد.

نتیجه‌گیری

در این مطالعه از مقادیر دبی جریان و مقادیر کیفی متناظر (هدایت الکتریکی، کل مواد جامد محلول و بی‌کربنات‌ها) در خصوص شبیه‌سازی مقادیر هدایت الکتریکی و TDS چند متغیره در ایستگاه‌های تپیک، دیزج و چهریق علیا در حوضه دریاچه ارومیه در دوره آماری ۹۹-۱۳۵۰ استفاده شد. جهت شبیه‌سازی مقادیر EC و TDS از مقادیر دبی جریان و بی‌کربنات‌ها استفاده شد. شبیه‌سازی با استفاده از دو مدل جنگل تصادفی و رگرسیون فرآیند گاوسی صورت گرفت. نتایج بررسی روند تغییرات مقادیر مورد بررسی نشان داد که روند تغییرات دبی جریان در تمامی ایستگاه‌های مورد بررسی کاهش می‌باشد. تغییرات کاهش در دو ایستگاه دیزج و چهریق علیا در سطح اعتماد ۵ درصد معنی‌دار می‌باشد. نتایج بررسی روند تغییرات مقادیر کیفی جریان نشان داد که کاهش دبی جریان در رودخانه‌های مورد بررسی سبب افزایش غلظت مقادیر کیفی شده است و روند تغییرات مقادیر کیفی در دوره آماری مورد بررسی در ایستگاه‌های مورد بررسی افزایشی بوده است.

نتایج شبیه‌سازی مقادیر EC با استفاده از مدل جنگل تصادفی در ایستگاه‌های مورد بررسی نشان داد که در هر دو مرحله آموزش و آزمایش کارایی بیش از ۹۸ درصد را ارائه کرد. با توجه به حدود اطمینان ۹۵ درصد نتایج نشان داد که به غیر از چندین مورد بیش تخمین و کم تخمینی، غالب مقادیر شبیه‌سازی شده EC در محدوده اطمینان قرار گرفته است. این موضوع در ایستگاه‌های تپیک و دیزج بهتر نمایان شده است. نتایج بررسی مقادیر RMSE در شبیه‌سازی مقادیر EC نشان داد که مدل جنگل تصادفی در مرحله آموزش میزان خطای $۶/۸۸$ ، $۴/۷۹$ و $۲/۸۰$ میکروموس بر سانتی‌متر به ترتیب در ایستگاه‌های چهریق علیا، دیزج و تپیک برآورد شده است که نسبت به دوره آزمایش این میزان خطا به ترتیب حدود $۳/۵۲$ ، $۴/۷۴$ و $۱/۸۳$ میکروموس بر سانتی‌متر برآورد شده است. مدل رگرسیون فرآیند گاوسی میزان خطای $۱۴/۰۱$ ، $۹/۱۵$ و $۴/۳۳$ را در مرحله آموزش و میزان خطای $۱۶/۰۳$ ، $۶/۴۵$ و $۲/۶۹$ را در مرحله آزمایش در خصوص شبیه‌سازی مقادیر هدایت الکتریکی با توجه به مقادیر دبی جریان و میزان بی‌کربنات‌ها به ترتیب برای ایستگاه‌های چهریق علیا، دیزج و تپیک ارائه کرد. این مقدار خطا نسبت به مدل جنگل تصادفی برای ایستگاه‌های چهریق علیا، دیزج و تپیک در مرحله آموزش به ترتیب ۱۰۴ ، ۹۱ و ۵۵ درصد و در مرحله آزمایش به ترتیب ۳۵۴ ، ۳۶ و ۴۷ درصد افزایش یافته است. در مورد شبیه‌سازی مقادیر TDS با توجه به

با توجه به شکل ۱۴-۱۴-چپ می‌توان مشاهده کرد که در تمامی ایستگاه‌های مورد بررسی میزان خطای مدل GPR بیشتر از مدل RF در شبیه‌سازی مقادیر هدایت الکتریکی در فاز آموزش می‌باشد. همچنین با توجه به این شکل می‌توان اختلاف زیاد بین مقادیر RMSE در مدل RF و GPR مشاهده کرد. مدل جنگل تصادفی برای استخراج هر چه بیشتر اطلاعات ویژگی‌ها یا الگوهای مفید از داده‌ها، می‌توان چندین دور محاسبات را برای به دست آوردن مقادیر متوسط بهینه انجام دهد. این موضوع می‌تواند سبب برتری این مدل باشد. در فاز آزمایش با توجه به شکل ۱۴-راست نیز می‌توان برتری مدل RF نسبت به مدل GPR را در تمامی ایستگاه‌ها در فاز آزمایش مشاهده کرد. در ایستگاه چهریق علیا در شکل ۱۴-راست تفاوت زیادی بین مقادیر RMSE در فاز آزمایش می‌باشد که نسبت به دو ایستگاه دیزج و تپیک بیشتر می‌باشد. به طور کلی می‌توان در این مطالعه دقت مدل RF را برتر از مدل GPR در شبیه‌سازی مقادیر هدایت الکتریکی در منطقه مورد مطالعه با توجه به مقادیر دبی جریان، کل مواد جامد محلول و بی‌کربنات‌ها معرفی کرد. شکل ۱۵-۱۵-چپ نتایج شبیه‌سازی مقادیر TDS در ایستگاه‌های مورد بررسی در منطقه مورد مطالعه در فاز آموزش با استفاده از دو مدل RF و GPR را نشان می‌دهد. با توجه به شکل ۱۵-چپ می‌توان مشاهده کرد که در ایستگاه چهریق علیا میزان خطای مدل GPR بیشتر از مدل RF می‌باشد. در ایستگاه دیزج تفاوت زیادی بین دو مدل RF و GPR از نظر آماره RMSE در فاز آموزش مشاهده نشد. اما در خصوص ایستگاه تپیک در فاز آموزش تا حدودی میزان خطای مدل RF بیشتر از مدل GPR برآورد شد. اما در فاز آزمایش، نتایج شبیه‌سازی TDS در منطقه مورد مطالعه با توجه به شکل ۱۵-راست نشان داد که در مورد ایستگاه دیزج همانند فاز آموزش تفاوتی بین دو مدل RF و GPR وجود ندارد. اما در ایستگاه چهریق علیا میزان خطای مدل RF کمتر از GPR می‌باشد که این موضوع در مورد ایستگاه تپیک این موضوع برعکس بوده و میزان خطای مدل GPR کمتر از مدل RF است. (Shabani et al (2020) در تحقیقات خود در شبیه‌سازی مقادیر تبخیر از تشت نشان دادند که عملکرد هر سه مدل جنگل تصادفی، رگرسیون فرآیند گاوسی و رگرسیون بردار پشتیبان عملکرد مناسبی در ایستگاه‌های مختلف داشتند. نتایج تحقیق حاضر با نتایج تحقیقات (Shabani et al (2020) مطابقت دارد. Hultquist et al (2014) نیز در تحقیقات خود مبنی بر شبیه‌سازی وسعت سوختگی در جنگل‌ها دقت و عملکرد و همچنین قطعیت دو مدل جنگل تصادفی و رگرسیون فرآیند گاوسی را مورد قبول ارزیابی کردند. نتایج تحقیق آن‌ها نشان داد که در زمینه‌های مختلف نیز عملکرد این دو مدل قابل قبول بوده و به نوع داده مورد بررسی وابسته نیست. این موضوع را می‌توان در مطالعات (Sun et al (2022) در مدل‌سازی چند متغیره

- Birgé, L. 2004. Model selection for Gaussian regression with random design. *Bernoulli*. 10(6):1039-1051.
- Breiman, L. 2001. Random Forests. *Machine Learning*. 45(1):5-32.
- Burt, D., Rasmussen, C. E. and Van Der Wilk, M. 2019. Rates of convergence for sparse variational Gaussian process regression. In *International Conference on Machine Learning*. PMLR: 862-871
- Chen, S. T., Yu, P. S. and Tang, Y. H. 2010. Statistical downscaling of daily precipitation using support vector machines and multivariate analysis. *Journal of hydrology*. 385(1-4):13-22.
- Chu, J. L., Kang, H., Tam, C. Y., Park, C. K. and Chen, C. T. 2008. Seasonal forecast for local precipitation over northern Taiwan using statistical downscaling. *Journal of Geophysical Research: Atmospheres*. 113(D12).
- Dibike, Y. B., Velickov, S., Solomatine, D. and Abbott, M. B. 2001. Model induction with support vector machines: introduction and applications. *Journal of Computing in Civil Engineering*. 15(3):208-216.
- Duan, W., He, B., Nover, D., Yang, G., Chen, W., Meng, H. and Liu, C. 2016. Water quality assessment and pollution source identification of the eastern Poyang Lake Basin using multivariate statistical methods. *Sustainability*. 8(2):133.
- Emamgholizadeh, S., Kashi, H., Marofpoor, I. and Zalaghi, E. 2014. Prediction of water quality parameters of Karoon River (Iran) by artificial intelligence-based models. *International Journal of Environmental Science and Technology*. 11(3):645-656.
- Eslami, P., Nasirian, A., Akbarpour, A. and Nazeri Tahroudi, M. 2022. Groundwater estimation of Ghayen plain with regression-based and hybrid time series models. *Paddy and Water Environment*. 1-12.
- Friedman, J., Hastie, T. and Tibshirani, R. 2001. *The elements of statistical learning* (Vol. 1, No. 10). New York: Springer series in statistics.
- George, J., Janaki, L. and Parameswaran Gomathy, J. 2016. Statistical downscaling using local polynomial regression for rainfall predictions—a case study. *Water resources management*. 30(1):183-193.
- Ghosh, S. 2010. SVM-PGSL coupled approach for statistical downscaling to predict rainfall from GCM output. *Journal of Geophysical Research: Atmospheres*. 115(D22).
- مقادیر دبی جریان و بی‌کربنات‌ها نتایج مقایسه دو مدل جنگل تصادفی و رگرسیون فرآیند گاوسی نشان داد که در ایستگاه چهریق علیا، مدل جنگل تصادفی نتایج بهتری نسبت به مدل رگرسیون فرآیند گاوسی ارائه کرده است که این برتری در مرحله آموزش و آزمایش به ترتیب حدود ۱۲۱ و ۲۶۷ درصد می‌باشد. اما در دو ایستگاه دیزج و تپیک، مدل رگرسیون فرآیند گاوسی نتایج بهتری نسبت به مدل جنگل تصادفی ارائه کرده است. نتایج مقادیر RMSE نشان داد که در مرحله آموزش تفاوت معنی‌داری بین دو مدل جنگل تصادفی و رگرسیون فرآیند گاوسی وجود ندارد. اما در مرحله آزمایش، نتایج شبیه‌سازی‌ها برتری کامل مدل رگرسیون فرآیند گاوسی را نسبت به مدل جنگل تصادفی به میزان ۶۵ و ۷۳ درصد به ترتیب در ایستگاه‌های دیزج و تپیک نشان داد. با توجه به همبستگی مناسب بین مقادیر دبی جریان و بی‌کربنات‌ها با مقادیر هدایت الکتریکی، دخالت این پارامترها مقادیر شبیه‌سازی شده را با قطعیت بالاتری ارائه می‌کند. برتری مدل جنگل تصادفی در شبیه‌سازی مقادیر EC در همه ایستگاه‌ها و مقادیر TDS در ایستگاه چهریق علیا تأیید و مشاهده شد.
- ### منابع
- Abaurrea, J. and Asín, J. 2005. Forecasting local daily precipitation patterns in a climate change scenario. *Climate Research*. 28(3):183-197.
- Adamowski, J. and Sun, K. 2010. Development of a coupled wavelet transform and neural network method for flow forecasting of non-perennial rivers in semi-arid watersheds. *Journal of Hydrology*. 390(1-2):85-91.
- Ahmadi, F., Nazeri Tahroudi, M., Mirabbasi, R., Khalili, K. and Jhajharia, D. 2018. Spatiotemporal trend and abrupt change analysis of temperature in Iran. *Meteorological Applications*. 25(2):314-321.
- Ahmadi, F., Nazeri Tahroudi, M., Mirabbasi, R., Kumar, R. 2022. Spatiotemporal analysis of precipitation and temperature concentration using PCI and TCI: a case study of Khuzestan Province, Iran. *Theor Appl Climatol*, <https://doi.org/10.1007/s00704-022-04077-6>.
- Alqahtani, A., Shah, M. I., Aldrees, A. and Javed, M. F. 2022. Comparative Assessment of Individual and Ensemble Machine Learning Models for Efficient Analysis of River Water Quality. *Sustainability*. 14(3):1183.
- Bhagwat, P. P. and Maity, R. 2012. Multistep-ahead river flow prediction using LS-SVR at daily scale. *Journal of water Resource and Protection*. 4(07):528.

- Kisi, O. 2012. Modeling discharge-suspended sediment relationship using least square support vector machine. *Journal of hydrology*. 456:110-120.
- Kumar S., Merwade V., Kam J., and Thurner K. 2009. Streamflow trends in Indiana: Effects of long term persistence, precipitation and subsurface drains. *Journal of Hydrology*. 374(1-2): 171-183.
- Kundu, S., Khare, D. and Mondal, A. 2017. Future changes in rainfall, temperature and reference evapotranspiration in the central India by least square support vector machine. *Geoscience Frontiers*. 8(3):583-596.
- Landman, W. A. and Mason, S. J. 2001. Forecasts of near-global sea surface temperatures using canonical correlation analysis. *Journal of Climate*. 14(18):3819-3833.
- Liong, S. Y. and Sivapragasam, C. 2002. Flood stage forecasting with support vector machines 1. *JAWRA Journal of the American Water Resources Association*. 38(1):173-186.
- Mandal, S., Srivastav, R. K. and Simonovic, S. P. 2016. Use of beta regression for statistical downscaling of precipitation in the Campbell River basin, British Columbia, Canada. *Journal of Hydrology*. 538:49-62.
- Mann, H.B. 1945. Non-parametric tests against trend. *Econometrica*, 13, MathSci Net, 245-259.
- Misra, D., Oommen, T., Agarwal, A., Mishra, S. K. and Thompson, A. M. 2009. Application and analysis of support vector machine based simulation for runoff and sediment yield. *Biosystems engineering*. 103(4):527-535.
- Montgomery, D. C., Peck, E. A. and Vining, G. G. 2021. *Introduction to linear regression analysis*. John Wiley & Sons.
- Najah, A., El-Shafie, A., Karim, O. A. and El-Shafie, A. H. 2013. Application of artificial neural networks for water quality prediction. *Neural Computing and Applications*. 22(1):187-201.
- Nazeri Tahroudi, M., Khalili, K., Ahmadi, F., Mirabbasi, R. and Jhajharia, D. 2019. Development and application of a new index for analyzing temperature concentration for Iran's climate. *International Journal of Environmental Science and Technology*. 16(6):2693-2706.
- Nazeri-Tahroudi, M. and Ramezani, Y. 2020. Estimation of Dew Point Temperature in Different
- He, Z., Wen, X., Liu, H. and Du, J. 2014. A comparative study of artificial neural network, adaptive neuro fuzzy inference system and support vector machine for forecasting river flow in the semiarid mountain region. *Journal of Hydrology*. 509:379-386.
- Hultquist, C., Chen, G. and Zhao, K. 2014. A comparison of Gaussian process regression, random forests and support vector regression for burn severity assessment in diseased forests. *Remote sensing letters*. 5(8):723-732.
- Iannace, G. and Ciaburro, G. 2021. Modelling sound absorption properties for recycled polyethylene terephthalate-based material using Gaussian regression. *Building Acoustics*. 28(2):185-196.
- Iannace, G., Ciaburro, G. and Trematerra, A. 2018. Heating, ventilation, and air conditioning (HVAC) noise detection in open-plan offices using recursive partitioning. *Buildings*. 8(12):169.
- Iannace, G., Ciaburro, G. and Trematerra, A. 2019a. Fault diagnosis for UAV blades using artificial neural network. *Robotics*. 8(3):59.
- Iannace, G., Ciaburro, G. and Trematerra, A. 2019b. Wind turbine noise prediction using random forest regression. *Machines*. 7(4):69.
- Khalili, K., Tahoudi, M. N., Mirabbasi, R. and Ahmadi, F. 2016. Investigation of spatial and temporal variability of precipitation in Iran over the last half century. *Stochastic Environmental Research and Risk Assessment*. 30(4):1205-1221.
- Khalili M.N., Ouarda T.B.M.J., and Gachon P. 2009. Identification of temporal trends in annual and seasonal low flows occurring in Canadian rivers: The effect of short- and long-term persistence, *Journal of Hydrology*. 369:183-197.
- Khozeymehnezhad, H. and Tahroudi, M. N. 2019. Annual and seasonal distribution pattern of rainfall in Iran and neighboring regions. *Arabian Journal of Geosciences*. 12(8):1-11.
- Kim, G. B., Kim, W. J., Kim, H. U. and Lee, S. Y. 2020. Machine learning applications in systems metabolic engineering. *Current opinion in biotechnology*. 64:1-9.
- Kim, Y. J. and Gu, C. 2004. Smoothing spline Gaussian regression: more scalable computation via efficient approximation. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*. 66(2):337-356.

- and Jhajharia, D. 2019. Development and application of a new index for analyzing temperature concentration for Iran's climate. *International Journal of Environmental Science and Technology*. 16(6):2693-2706.
- Tripathi, S., Srinivas, V. V. and Nanjundiah, R. S. 2006. Downscaling of precipitation for climate change scenarios: a support vector machine approach. *Journal of hydrology*. 330(3-4):621-640.
- Tung, T. M. and Yaseen, Z. M. 2021. Deep learning for prediction of water quality index classification: tropical catchment environmental assessment. *Natural Resources Research*. 30(6):4235-4254.
- Virmani, C., Choudhary, T., Pillai, A. and Rani, M. 2020. Applications of machine learning in cyber security. In *Handbook of research on machine and deep learning applications for cyber security* (pp. 83-103). IGI Global.
- Wang, F., Wang, Y., Zhang, K., Hu, M., Weng, Q. and Zhang, H. 2021. Spatial heterogeneity modeling of water quality based on random forest regression and model interpretation. *Environmental Research*. 202, 111660, <https://doi.org/10.1016/j.envres.2021.111660>
- Wang, Q. Fan, X. Qin, Z. Wang, M. 2012. Change trends of temperature and precipitation in the Loess Plateau Region of China, 1961–2010. *Global and Planetary Change*. 93:138-147.
- Wang, W. C., Xu, D. M., Chau, K. W. and Lei, G. J. 2014. Assessment of river water quality based on theory of variable fuzzy sets and fuzzy binary comparison method. *Water resources management*. 28(12):4183-4200.
- Wang, W., Van Gelder, P. H. A. J. M., Vrijling, J. K., and Ma, J. 2005. Testing and modeling autoregressive conditional heteroskedasticity of streamflow processes. *Nonlinear processes in Geophysics*. 12:1. 55-66.
- Weisberg, S. 2005. *Applied linear regression* (Vol. 528). John Wiley & Sons.
- Wilby, R. L., Dawson, C. W. and Barrow, E. M. 2002. SDSM—a decision support tool for the assessment of regional climate change impacts. *Environmental Modelling & Software*. 17(2):145-157.
- Yoon, H., Jun, S. C., Hyun, Y., Bae, G. O. and Lee, K. K. 2011. A comparative study of artificial neural networks and support vector machines for predicting Climates of Iran Using Support Vector Regression. *IDŐJÁRÁS/QUARTERLY JOURNAL OF THE HUNGARIAN METEOROLOGICAL SERVICE*. 124(4):521-539.
- Raje, D. and Mujumdar, P. P. 2011. A comparison of three methods for downscaling daily precipitation in the Punjab region. *Hydrological Processes*. 25(23):3575-3589.
- Raji, M., Tahroudi, M. N., Ye, F. and Dutta, J. 2022. Prediction of heterogeneous Fenton process in treatment of melanoidin-containing wastewater using data-based models. *Journal of Environmental Management*. 307:114518.
- Ramezani, Y., Khashei-Siuki, A. and Nazeri Tahroudi, M. 2020. Spatial distribution of the daily, monthly, and annual precipitation concentration indices in the Lake Urmia basin, Iran. *Quarterly journal of the Hungarian meteorological service*. 124(1):73-95.
- Saha, A., Tso, S., Rabski, J., Sadeghian, A. and Cusimano, M. D. 2020. Machine learning applications in imaging analysis for patients with pituitary tumors: a review of the current literature and future directions. *Pituitary*. 23(3):273-293.
- Sain, S. R., Baggerly, K. A. and Scott, D. W. 1994. Cross-validation of multivariate densities. *Journal of the American Statistical Association*. 89(427):807-817.
- Shabani, S., Samadianfard, S., Sattari, M. T., Mosavi, A., Shamshirband, S., Kmet, T. and Várkonyi-Kóczy, A. R. 2020. Modeling pan evaporation using Gaussian process regression K-nearest neighbors random forest and support vector machines; comparative analysis. *Atmosphere*. 11(1):66.
- Singh, K. P., Basant, A., Malik, A. and Jain, G. 2009. Artificial neural network modeling of the river water quality—a case study. *Ecological modelling*. 220(6):888-895.
- Sun, N., Zhang, S., Peng, T., Zhang, N., Zhou, J. and Zhang, H. 2022. Multi-Variables-Driven Model Based on Random Forest and Gaussian Process Regression for Monthly Streamflow Forecasting. *Water*. 14(11):1828.
- Tabatabaei, S. M., Tahroudi, M. N. and Hamraz, B. S. 2021. Comparison of the performances of GEP, ANFIS, and SVM artificial intelligence models in rainfall simulaton. *Quarterly Journal of the Hungarian Meteorological Service*. 125(2):195-209.
- Tahroudi, M. N., Khalili, K., Ahmadi, F., Mirabbasi, R.

July. Stochastic variational inference for Bayesian sparse Gaussian process regression. In 2019 International Joint Conference on Neural Networks (IJCNN) (pp. 1-8). IEEE.

groundwater levels in a coastal aquifer. Journal of hydrology. 396(1-2):128-138.

Yu, H., Nghia, T., Low, B. K. H. and Jaillet, P. 2019.

Prediction of TDS and EC Values of River Using Machine Learning Methods

M.H. Jabbar Matoori¹, A. Fathi², F. Ahmadi^{3*}

Received: Aug.25, 2022

Accepted: Sep.18, 2022

Abstract

In this study, while examining the performance of random forest model and Gaussian process regression, simulating the values of electrical conductivity and total dissolved solids in the western stations of Lake Urmia (Chehriq-Olya, Dizj and Tepik) according to the values of flow discharge and The HCO₃ of the river flow has been studied in the period of 1971-2020. According to the RMSE and NSE statistics, the simulation results of EC values in the studied stations showed that the error rate of the RF model is lower than the GPR model and the efficiency of the model is also higher. The error rate of simulating EC values using the RF model in the test phase in Chihriq-Olya, Dizj and Tepik stations is about 35%, 36 and 47% less than the GPR model. In general, the results showed that according to the simulated confidence intervals of EC and TDS parameters, the performance of the two investigated models is acceptable, but in the case of TDS values, the behavior of the two investigated models in the two stations of Dizj and Tepik is different.

Keywords: Gaussian Process, Lake Urmia, Random Forest, Regression, Water Quality

1- MS.c Student, Faculty of Water and Environmental Engineering, Shahid Chamran University of Ahvaz, Ahvaz, Iran

2 - Assistant Professor, Department of Environmental Engineering, Faculty of Water and Environmental Engineering, Shahid Chamran University of Ahvaz, Ahvaz, Iran

3- Assistant Professor, Department of Hydrology and Water Resources, Faculty of Water and Environmental Engineering, Shahid Chamran University of Ahvaz, Ahvaz, Iran

(* - Corresponding Author Email: f.ahmadi@scu.ac.ir)