

بررسی عملکرد الگوریتم جنگل تصادفی در پیش‌بینی نوسانات سطح ایستابی در مقایسه با دو مدل درخت تصمیم و شبکه عصبی مصنوعی (مطالعه موردی: آبخوان آزاد دشت بیرجند)

فاطمه پورصالحی^۱، عباس خاشعی سیوکی^{۲*}، سید رضا هاشمی^۳

۱. دانشجوی دکترای مهندسی منابع آب، دانشکده کشاورزی، دانشگاه بیرجند، بیرجند، ایران

۲. استاد گروه علوم و مهندسی آب، دانشکده کشاورزی، دانشگاه بیرجند، بیرجند، ایران

۳. دانشیار گروه علوم و مهندسی آب، دانشکده کشاورزی، دانشگاه بیرجند، بیرجند، ایران

(تاریخ دریافت ۱۴۰۰/۰۴/۲۴؛ تاریخ تصویب ۱۴۰۰/۱۰/۲۱)

چکیده

امروزه به سبب برداشت‌های بی‌رویه از منابع آب زیرزمینی و افت تراز سطح ایستابی به‌ویژه در مناطق خشک و نیمه‌خشک، برنامه‌ریزی و مدیریت در مصرف این منابع با ارزش اهمیت زیادی دارد که این امر نیازمند مطالعه رفتار آبخوان نسبت به تغییرات وارد شده بر آن است. هدف از انجام این تحقیق، بررسی کارایی الگوریتم جنگل تصادفی در پیش‌بینی تراز سطح ایستابی آبخوان آزاد دشت بیرجند و مقایسه نتایج آن با دو مدل درخت تصمیم و شبکه عصبی مصنوعی است. در این راستا، ابتدا اطلاعات ورودی به مدل شامل تراز سطح ایستابی چاه‌های مشاهده‌ای، دما، بارندگی، رطوبت و تبخیر طی سال‌های آبی ۱۳۸۹-۱۳۹۰ تا ۱۳۹۵-۱۳۹۶ به صورت ماهانه جمع‌آوری و پس از بررسی روند و حذف آن، برای ایجاد مدل‌های یادشده از بسته نرم‌افزاری Rattle در نرم‌افزار آماری R استفاده شد. نتایج حاصل از شبیه‌سازی با استفاده از الگوریتم جنگل تصادفی براساس معیارهای ارزیابی معادل $R^2=0/714$ ، $RMSE=0/03$ متر و $NS=0/598$ نشان می‌دهد این الگوریتم توانایی نسبتاً زیادی در شبیه‌سازی تراز سطح ایستابی آبخوان دارد. از مقایسه نتایج این الگوریتم با دو مدل درخت تصمیم و شبکه عصبی مصنوعی می‌توان دریافت که نتایج الگوریتم جنگل تصادفی نسبت به مدل درخت تصمیم با $R^2=0/549$ ، $RMSE=0/072$ متر و $NS=-0/187$ تطابق بیشتری با تراز واقعی آبخوان دارد و با نتایج شبکه عصبی مصنوعی با $R^2=0/7055$ ، $RMSE=0/003$ متر و $NS=0/6046$ هم‌راستا است. همچنین، خروجی الگوریتم جنگل تصادفی نشان می‌دهد در بین پارامترهای ورودی، چاه‌های مشاهده‌ای واقع در نواحی مرکزی دشت و نیز پارامترهای هواشناسی بارندگی و رطوبت در شبیه‌سازی تراز سطح ایستابی آبخوان نقش مؤثرتری نسبت به سایر پارامترها ایفا می‌کنند.

واژه‌های کلیدی: آب زیرزمینی، شبیه‌سازی تراز سطح ایستابی، معیارهای ارزیابی، نرم‌افزار آماری R.

مقدمه

بهره‌برداری بی‌رویه از منابع آب زیرزمینی در سال‌های اخیر سبب برهم زدن تعادل این منابع شده و افت آب زیرزمینی در آبخوان‌های بسیاری از نقاط کشور افزایش یافته است. به منظور آگاهی از وضعیت این منابع و مدیریت بهینه آن‌ها، باید پیش‌بینی دقیقی از نوسان‌های تراز آب زیرزمینی صورت گیرد [۱]. از این‌رو، یکی از مباحث مهم در مدیریت حوضه آبخیز، پیش‌بینی تراز سطح آب زیرزمینی است تا بتوان با توجه به آن، برنامه‌ریزی‌های لازم در این زمینه را انجام داد و مدیریت موفق منابع آب در حوضه آبخیز نیازمند رویکردهای جهت‌دار، جامع و هدفمند است تا بتواند با وجود روند تشدید مشکلات مربوط به آب و افزایش رو به رشد تقاضای آن، نیاز مصرف‌کنندگان را برطرف سازد [۲]. مدل‌سازی آب‌های زیرزمینی تا کنون با ابزارهای مختلفی انجام شده است. اگرچه مدل‌های ریاضی و فیزیکی ابزار اصلی برای بررسی و شناخت فرایندهای فیزیکی حاکم بر یک سیستم هستند، اما نبود اطلاعات و داده‌های مورد نیاز این مدل‌ها، آن‌ها را با محدودیت کاربرد روبه‌رو کرده است. این مسئله باعث روی آوردن به مدل‌های هوشمند به عنوان جایگزین مدل‌های ریاضی و فیزیکی شده است [۳ و ۴]. از طرفی در سال‌های اخیر، رشد روزافزون کاربرد هوش محاسباتی در حل مسائلی مشاهده می‌شود که تأثیر پارامترهای بسیار زیاد در برخی فرایندهای فیزیکی و روابط کاملاً غیرخطی میان آن‌ها حاکم است و مبنای روش‌های هوشمند، استفاده از دانش نهفته در داده‌ها، تلاش برای استخراج روابط ذاتی بین آن‌ها و تعمیم آن در موقعیت‌های دیگر است [۲].

در دهه‌های اخیر، روش‌های پیش‌بینی متعددی، شامل مدل‌های هیدرودینامیکی مبتنی بر فیزیک (مانند: CHAM، MIKE21 و EFDC)، تجزیه و تحلیل سری‌های زمانی (مانند: میانگین متحرک خودهمبسته و میانگین متحرک خودهمبسته تجمعی)، روش‌های محاسباتی نرم (مانند: شبکه عصبی مصنوعی (ANN))، رگرسیون بردار پشتیبان (SVR) و مدل‌های درختی برای شبیه‌سازی سری‌های زمانی هیدرولوژیکی توسعه یافته‌اند [۵-۷]. در این راستا، خلیلی‌نفت‌چالی و همکاران (۲۰۱۷)،

توانایی الگوریتم‌های تنبیل^۱ KNN، KSTAR و LWL و نیز مدل درختی M5 در برآورد تراز سطح ایستابی آبخوان دشت نیشابور را مورد ارزیابی قرار دادند. نتایج پژوهش آنها نشان داد عملکرد مدل‌های KNN، KSTAR و M5 بسیار به هم نزدیک است، در حالی که مدل LWL نسبت به سایر مدل‌ها ضعیف‌تر عمل می‌کند [۸]. میرهاشمی و همکاران (۲۰۱۸) با استفاده از الگوریتم درختی CART نوسان‌های سطح آب زیرزمینی در داخل و خارج شبکه تحت آبیاری شهرستان قزوین را شبیه‌سازی کردند. نتایج پژوهش آنها نشان داد این الگوریتم در داخل شبکه آبیاری کارایی بیشتری در شبیه‌سازی نوسانات سطح ایستابی نسبت به خارج از آن محدوده دارد [۹].

محتشم و همکاران (۲۰۱۰) با استفاده از روش شبکه عصبی مصنوعی سطح ایستابی دشت بیرجند را پیش‌بینی کردند. نتایج این تحقیق نشان داد با استفاده از این روش و تدقیق تخمین میزان آب برداشتی می‌توان سطح آب زیرزمینی را تا ۱۲ ماه بعد و با دقت بیشتر نسبت به روش‌هایی که از اطلاعات تبخیر از سطح گیاه مرجع به عنوان شاخص برداشت آب استفاده می‌کنند، تخمین زد [۱۰]. خاشعی‌سیوکی و همکاران (۲۰۱۳) مدل‌های شبکه عصبی مصنوعی، ANFIS و رگرسیونی را برای برآورد سطح ایستابی آبخوان دشت نیشابور با یکدیگر مقایسه کردند. نتایج پژوهش آنها نشان داد مدل شبکه عصبی مصنوعی در ماه آذر و مدل ANFIS در ماه شهریور بهترین کارایی را دارند، در حالی که مدل رگرسیونی کمترین دقت را دارد [۱۱]. رجائی و ابراهیمی (۲۰۱۶) با استفاده از مدل ترکیبی موجک شبکه عصبی پویا عمق آب زیرزمینی را پیش‌بینی و نتایج به‌دست‌آمده را با مدل‌های شبکه عصبی و رگرسیون خطی چندمتغیره مقایسه کردند. نتایج پژوهش آنها نشان داد مدل ترکیبی موجک شبکه عصبی نسبت به دو مدل دیگر در مدل‌سازی نوسان‌های ماهانه آب زیرزمینی توانایی زیادی دارد [۴]. سعیدی رضوی و عرب (۲۰۱۹) سطح آب زیرزمینی دشت عجب‌شیر را با استفاده از مدل‌های منطق فازی، شبکه عصبی و سری زمانی پیش‌بینی کردند. نتایج تحقیق بر دقت زیاد مدل منطق فازی نسبت به سایر روش‌ها دلالت دارد [۱۲].

فازی، جنگل تصادفی و شبکه عصبی پیش‌بینی کردند. نتایج پژوهش یادشده نشان داد مدل جنگل تصادفی دقت زیادی در پیش‌بینی تراز سطح ایستابی دارد [۱۸]. وانگ و همکاران (۲۰۱۸) به پیش‌بینی کوتاه‌مدت تراز آب زیرزمینی با استفاده از رگرسیون بهبودیافته جنگل تصادفی با ترکیبی از ویژگی‌های تصادفی پرداختند. نتایج پژوهش یادشده بر دقت زیاد این روش در پیش‌بینی تغییرات در سطح آب زیرزمینی دلالت دارد [۱۹].

با توجه به فرارگیری دشت بیرجند در اقلیم خشک و نیمه‌خشک ضروری است با به‌کارگیری روش‌های محاسباتی صحیح و مدل‌های هوشمند، تراز سطح ایستابی آبخوان پیش‌بینی شود تا با شبیه‌سازی رفتار آبخوان از بروز آسیب‌های احتمالی در آینده که بر اثر افت تراز سطح ایستابی رخ می‌دهد، جلوگیری به عمل آید. هدف از انجام این مطالعه، ارزیابی عملکرد الگوریتم جنگل تصادفی در پیش‌بینی تراز آبخوان آزاد دشت بیرجند و نیز بررسی دقت این روش در مقایسه با مدل‌های درخت تصمیم و شبکه عصبی مصنوعی است که براساس بررسی‌های انجام‌شده، تا کنون مطالعه‌ای برای مقایسه عملکرد سه مدل یادشده در شبیه‌سازی تراز سطح ایستابی در این دشت صورت نگرفته است.

مواد و روش‌ها

منطقه پژوهش

دشت بیرجند در قسمت شمالی ارتفاعات باقران با مختصات تقریبی ۳۲ درجه و ۳۴ دقیقه تا ۳۳ درجه و ۸ دقیقه عرض شمالی و ۵۸ درجه و ۴۱ دقیقه تا ۵۹ درجه و ۴۴ دقیقه طول شرقی قرار گرفته است. حوضه آبریز بیرجند دارای وسعت ۳۱۵۵ کیلومتر مربع بوده، که ۱۸۴۵ کیلومتر دشت و بقیه را ارتفاعات تشکیل می‌دهد. این دشت حالت کشیده دارد و تمامی پیرامون آن را ارتفاعات و بخش مرکزی را آبخوان آبرفتی تشکیل می‌دهد. دشت بیرجند طبق طبقه‌بندی‌های اقلیمی جزء مناطق خشک محسوب می‌شود. از نظر توپوگرافی حداکثر ارتفاع حوضه از سطح دریا ۲۷۸۷ متر در ارتفاعات شمالی منطقه (کوه بند دره) و حداقل ارتفاع نیز ۱۱۸۰ متر در خروجی دشت (منطقه فدشک) است [۲۰]. شکل ۱ موقعیت مکانی آبخوان دشت بیرجند را نشان می‌دهد.

جبال‌بارزی و ملکیان (۲۰۱۹)، کارایی دو روش شبکه عصبی مصنوعی و برنامه‌ریزی بیان ژن را در پیش‌بینی سطح ایستابی دشت جیرفت مقایسه کردند. نتایج پژوهش یادشده دقت زیاد روش شبکه عصبی مصنوعی در پراکنش داده‌های دشت جیرفت را نشان می‌دهد [۱۳]. نایاک و همکاران (۲۰۰۶) برای پیش‌بینی تراز سطح ایستابی در آبخوان ساحلی هند از مدل شبکه عصبی مصنوعی بهره بردند. نتایج پژوهش آنها نشان داد این مدل‌ها توانایی پیش‌بینی سطح ایستابی به مدت چهار ماه را دارند [۱۴]. سربکانت و همکاران (۲۰۰۹) کاربرد مدل شبکه عصبی در پیش‌بینی نوسان‌های زمانی سطح آب زیرزمینی را در حوضه آبخیز مهشوارم هندوستان بررسی کردند. نتایج پژوهش یادشده نشان‌دهنده توانایی این روش در مدل‌سازی تراز آبخوان است [۱۵]. سان و همکاران (۲۰۱۶) کارایی شبکه عصبی مصنوعی را در پیش‌بینی تراز آب زیرزمینی ارزیابی کردند. نتایج پژوهش آنها نشان داد روش ANN در مدت زمان یک روز پیش‌بینی دقیقی ارائه می‌دهد، در حالی که با افزایش زمان به مدت ۳ و ۷ روز از کارایی آن کاسته می‌شود [۱۶].

الگوریتم جنگل تصادفی (RF)^۱ یک روش گروهی است که چند الگوریتم درختی را برای تولید یک پیش‌بینی مکرر از هر پدیده ترکیب می‌کند. جنگل تصادفی می‌تواند الگوهای پیچیده را یاد بگیرد و ارتباط غیرخطی بین متغیرهای توضیحی و متغیرهای وابسته را در نظر بگیرد. همچنین، می‌تواند انواع مختلف داده‌ها را در تجزیه و تحلیل بگنجانند و ترکیب کند که این هم به علت نبود توزیع پیش‌فرض‌ها (توزیع نرمال) درباره داده‌های استفاده‌شده است. RF هزاران متغیر ورودی را بدون حذف یکی از آنها پذیرا است و اجرا می‌کند، همچنین می‌تواند برآورد کند که کدام متغیر در پیش‌بینی مدل مهم است. RF نسبت به شبکه‌های عصبی مصنوعی در شرایط گیر افتادن در کمینه محلی و داده‌های نویزدار، حساسیت کمتری دارد و می‌تواند تخمین بهتری از پارامترها را داشته باشد [۱۷].

در زمینه پیش‌بینی تراز سطح ایستابی با استفاده از روش جنگل تصادفی، نوروزی و ندیری (۲۰۱۸) سطح آب زیرزمینی دشت بوکان را با استفاده از سه مدل منطق



شکل ۱. موقعیت آبخوان دشت بیرجند در ایران و استان خراسان جنوبی [۲۱]

- ۱- نمونه‌های مختلف بوت استرپ X_i (i = تکرار بوت استرپ) به طور تصادفی از مجموعه داده اصلی X انتخاب می‌شوند. دو سوم از نمونه‌ها در یک نمونه بوت استرپ (InBag) و یک سوم از نمونه‌ها خارج از کیسه (OOB) گنجانده می‌شوند. هر درخت برای مطابقت با یک زیرمجموعه خاصی از بوت استرپ ساخته می‌شود.
- ۲- در یک گره در هر درخت، تقسیم جدید به طور تصادفی از همه شاخ‌ها و متغیر ورودی با کمترین میانگین مربع خطا (MSE) به عنوان معیار تقسیم درخت رگرسیون انتخاب می‌شود.
- ۳- فرایند تقسیم داده‌ها در هر گره داخلی مطابق با گام‌های بالا تا زمانی که تمام درختان تصادفی رشد کنند و شرایط توقف حاصل شود، تکرار می‌شود.
- ۴- نتایج نهایی رگرسیون می‌تواند براساس رابطه ۱ محاسبه شود:

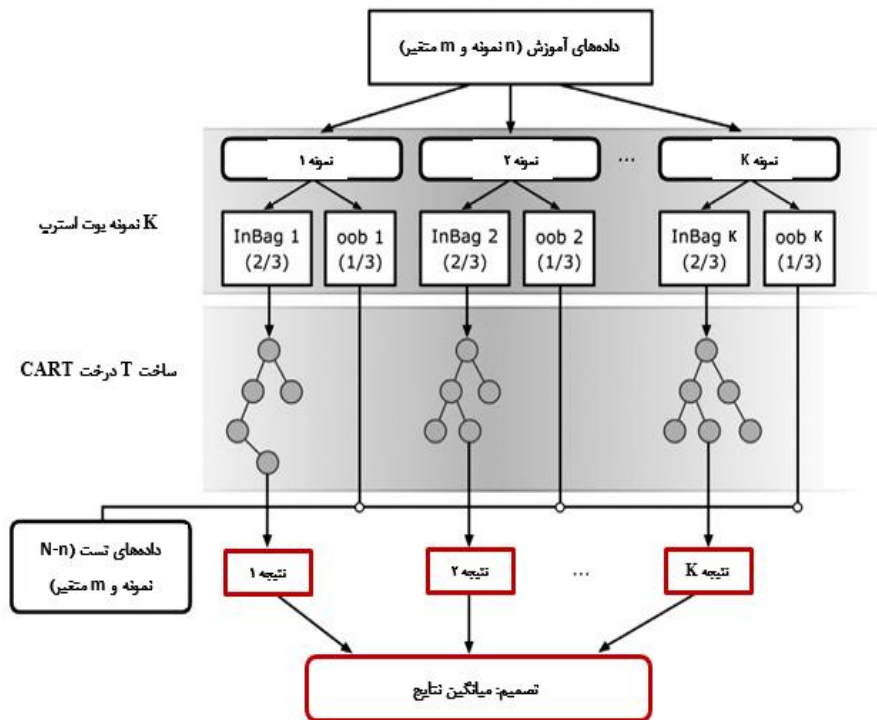
$$\hat{y}(x_i) = \frac{1}{K} \sum_{k=1}^K T_b(x_i) \quad (1)$$

در حالی که K تعداد درختان جنگل و T_b نشان‌دهنده هر درخت (CART: Classification and Regression Trees) است.

الگوریتم جنگل تصادفی

در الگوریتم RF برای تشکیل هر درخت، دسته متفاوتی از الگوهای موجود با در نظر گرفتن جایگزینی دوباره هر الگوی انتخاب‌شده، انتخاب می‌شوند. اندازه این دسته انتخاب‌شده، برابر تعداد کل الگوهای موجود خواهد بود. مدل RF در سال ۲۰۰۱ توسط برایمن [۲۲] به عنوان روشی از توسعه جدید درخت‌های تصمیم‌گیری، ارائه شد که پیش‌بینی چندین الگوریتم منفرد را با هم و با استفاده از قوانین مبتنی ترکیب می‌کند. اصول کلی تکنیک‌های آموزش گروهی بر پایه این فرض است که دقت آن‌ها از دیگر الگوریتم‌های آموزشی بیشتر است، زیرا ترکیبی از چند مدل پیش‌بینی، دقیق‌تر از یک مدل است و گروه‌ها در حالی که ضعف‌های طبقه‌ها را کاهش می‌دهند، هم‌زمان قدرت مجموعه‌های منفرد و منحصر به فرد طبقه‌ها را نیز بیشتر می‌کنند [۱۷ و ۲۳].

دو پارامتر تعریف‌شده توسط کاربر RF، شامل B ، تعداد درختان در جنگل و D ، تعداد ویژگی‌های مورد استفاده تقسیم گره‌ها می‌شود. مقدار پیش‌فرض B برابر ۵۰۰ است، اما با توجه به نوع کاربرد دوباره تنظیم می‌شود. یک سوم متغیرها برای انجام رگرسیون به مقدار پیش‌فرض D تنظیم می‌شوند. گام‌های الگوریتم رگرسیونی جنگل تصادفی (شکل ۲) به صورت زیر است [۱۹ و ۲۲]:



شکل ۲. نمودار گردش الگوریتم جنگل تصادفی [۲۴ و ۲۵]

پیش‌بینی‌کننده انجام می‌پذیرد، بازه‌های انشعاب طوری انتخاب می‌شوند که مجموع مجذور انحراف از میانگین داده‌های هر گره را به حداقل برسانند [۲۷].

مدل شبکه عصبی مصنوعی (ANN)

شبکه پرسپترون چندلایه^۳ (MLP) نوعی از شبکه‌های عصبی مصنوعی است که بر مبنای یک واحد محاسباتی به نام پرسپترون ساخته می‌شود. یک پرسپترون برداری از ورودی‌ها با مقادیر حقیقی را می‌گیرد و یک ترکیب خطی از این ورودی‌ها را محاسبه می‌کند. اگر مقدار حاصل از یک مقدار آستانه بیشتر بود، خروجی پرسپترون برابر با ۱ و در غیر این صورت، معادل ۰ خواهد بود. این شبکه شامل یک لایه ورودی، یک یا چند لایه پنهان و یک لایه خروجی است. برای آموزش این شبکه، معمولاً از الگوریتم پس‌انتشار خطا (BP) استفاده می‌شود. طی آموزش شبکه MLP به کمک الگوریتم یادگیری BP، ابتدا محاسبات از ورودی شبکه به سوی خروجی شبکه انجام می‌شود و سپس، مقادیر خطای محاسبه‌شده به لایه‌های قبل انتشار می‌یابد و وزن‌ها تصحیح می‌شوند. ابتدا، محاسبه خروجی

درخت تصمیم‌گیری^۱

روش درخت تصمیم یک روش سلسله‌مراتبی یا چند مرحله‌ای است که در آن به صورت بازگشتی مجموعه داده‌ها به روش دودویی به تقسیمات فرعی و کوچک‌تر تقسیم می‌شوند تا زمانی که تقسیمات فرعی نهایی نتوانند بیشتر از آن تجزیه شوند. درختان تصمیم استقرایی مجموعه‌ای از داده‌های معلوم را می‌گیرد و یک درخت تصمیم را از آن استخراج می‌کند. سپس، درخت می‌تواند به صورت مجموعه قوانینی برای پیش‌بینی نتیجه ویژگی‌های معلوم استفاده شود [۲۶]. یک درخت تصمیم معمولاً از چهار بخش ریشه، شاخه، گره‌ها و برگ‌ها تشکیل شده است که در آن گره‌ها با دایره مشخص می‌شوند و شاخه‌ها نشان‌دهنده اتصال بین گره‌ها هستند. برای رسم آسان‌تر درخت تصمیم معمولاً آن را از چپ به راست یا از بالا به پایین رسم می‌کنند، به گونه‌ای که ریشه (گره اول) در بالا قرار می‌گیرد. انتهای یک زنجیره را برگ می‌نامند. هر گره مربوط به یک ویژگی معین است و شاخه‌ها به معنای بازه‌ای از مقادیر هستند که بخش‌های مختلف مجموعه مقادیر معلوم ویژگی‌ها را در اختیار کاربر قرار می‌دهند. عمل انشعاب توسط یکی از متغیرهای

2. Artificial Neural Network
3. Multilayer perceptron

1. Decision Tree

$$\beta = \text{Median} \left[\frac{x_j - x_i}{j - i} \right] (\forall j > i) \quad (3)$$

در رابطه یادشده x_i و x_j به ترتیب مقادیر مشاهداتی z ام و i ام را نشان می‌دهد. مقادیر مثبت β روند افزایشی و مقادیر منفی آن روند کاهشی در سری زمانی را نشان می‌دهد [۳۱].

پس از حذف روند در داده‌های سری زمانی یادشده، برای شبیه‌سازی تراز سطح ایستابی آبخوان دشت بیرجند با استفاده از الگوریتم جنگل تصادفی، درخت تصمیم و شبکه عصبی مصنوعی از بسته نرم‌افزاری Rattle که در نرم‌افزار آماری R تعبیه شده است، استفاده شد. برای معرفی داده‌های ورودی به مدل، اطلاعات جمع‌آوری شده به صورت فایل اکسل با پسوند CSV ایجاد و در مدل فراخوانی شدند. در این تحقیق ۷۰ درصد داده‌ها به عنوان مرحله آموزش (train) و ۳۰ درصد داده‌ها به عنوان مرحله آزمون (test) به مدل وارد شدند. همچنین، در این شبیه‌سازی تراز سطح ایستابی هر ماه براساس داده‌های ورودی ماه قبل انجام شد. جدول ۱ و شکل ۳ به ترتیب مختصات چاه‌های مشاهده‌ای و محل قرارگیری آن‌ها در سطح دشت را نشان می‌دهد.

به صورت لایه‌به‌لایه انجام می‌شود و خروجی هر لایه، ورودی لایه بعدی خواهد بود [۲۸].

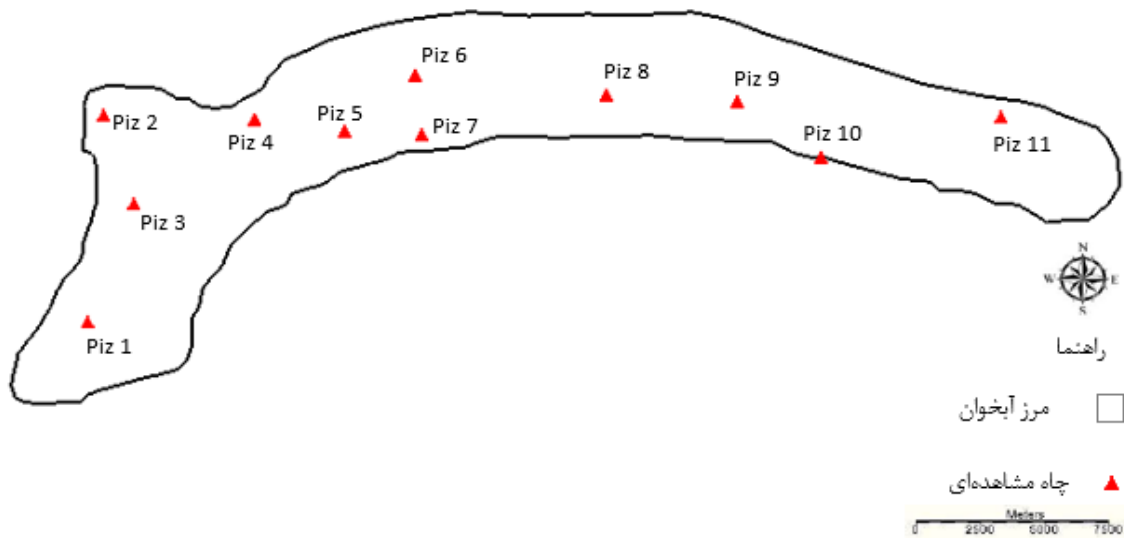
در این مطالعه ابتدا اطلاعات تراز سطح ایستابی ۱۱ پیزومتر (چاه‌های مشاهده‌ای) و پارامترهای هواشناسی شامل دما، بارندگی، رطوبت و تبخیر در بازه زمانی سال‌های آبی ۱۳۸۹-۱۳۹۰ تا ۱۳۹۵-۱۳۹۶ به صورت ماهانه جمع‌آوری و بانک اطلاعاتی مورد نیاز برای ایجاد تراز سطح ایستابی با استفاده از روش‌های مورد نظر تهیه شد. سپس، با استفاده از آزمون من-کندال فصلی در نرم‌افزار XLSTAT نسخه ۲۰۱۹ وجود روند در سری زمانی داده‌ها مورد بررسی قرار گرفت. پس از اثبات وجود روند، با استفاده از رابطه ۲ روند از سری زمانی داده‌ها حذف شد [۲۹ و ۳۰]:

$$D = D_i - i \times \beta \quad (2)$$

در این رابطه D تراز سطح ایستابی پس از حذف روند در هر ماه، D_i تراز سطح ایستابی قبل از حذف روند در هر ماه، i شماره ردیف هر ماه و β شیب خط روند تراز سطح ایستابی است. شیب خط روند سری زمانی در آزمون من-کندال براساس روش تخمین گر سن^۱ با استفاده از رابطه ناپارامتری ۳ به دست می‌آید:

جدول ۱. مختصات چاه‌های مشاهده‌ای

تراز سطح زمین (m)	UTMy	UTMx	چاه مشاهده‌ای
۱۳۰۰/۵۲	۳۶۲۹۳۶۱	۶۷۳۸۰۲	PIZ 1
۱۴۰۶/۶۴	۳۶۳۸۳۸۳	۶۷۴۴۹۹	PIZ 2
۱۳۵۱	۳۶۳۴۵۲۸	۶۷۵۷۹۵	PIZ 3
۱۳۶۵/۱۲	۳۶۳۸۲۲۷	۶۸۰۹۴۹	PIZ 4
۱۳۵۹	۳۶۳۷۶۹۵	۶۸۴۷۷۳	PIZ 5
۱۳۷۸/۷۱	۳۶۴۰۱۱۹	۶۸۷۷۹۰	PIZ 6
۱۳۹۲	۳۶۳۷۵۴۲	۶۸۸۰۵۸	PIZ 7
۱۴۲۱/۷۷	۳۶۳۹۲۷۳	۶۹۵۹۴۰	PIZ 8
۱۴۲۵/۴۴	۳۶۳۸۹۷۰	۷۰۱۵۳۴	PIZ 9
۱۵۴۲	۳۶۳۶۵۴۸	۷۰۵۱۰۶	PIZ 10
۱۴۹۶/۹۵	۳۶۳۸۳۰۰	۷۱۲۸۰۰	PIZ 11



شکل ۳. محل قرارگیری چاه‌های مشاهده‌ای در آبخوان دشت بیرجند

$$\text{Accuracy} = (TP + TN) / (TP + FN + FP + TN) \quad (7)$$

پارامترهای TP، FN، TN و FP به ترتیب عبارت‌اند از: نمونه عضو دسته مثبت باشد و عضو همین کلاس تشخیص داده شود (مثبت صحیح^۳)، نمونه عضو کلاس مثبت باشد و عضو کلاس منفی تشخیص داده شود (منفی کاذب^۴)، نمونه عضو کلاس منفی باشد و عضو همین کلاس تشخیص داده شود (منفی صحیح^۵) و نمونه عضو کلاس منفی باشد و عضو کلاس مثبت تشخیص داده شود (مثبت کاذب^۶).

نتایج و بحث

در این تحقیق، ابتدا پارامترهای مؤثر بر پیش‌بینی تراز سطح ایستابی شامل چاه‌های مشاهده‌ای، دما، بارندگی، رطوبت و تبخیر به عنوان ورودی به مدل‌های مورد نظر تهیه شد. سپس، فرض وجود روند در هر یک از پارامترها ارزیابی شد. با توجه به ماهانه بودن سری زمانی، برای بررسی روند در داده‌ها از آزمون من-کندال فصلی در نرم‌افزار XLSTAT 2019 استفاده شد. نتایج این آزمون نشان داد در پارامترهای دما، بارندگی، رطوبت و تبخیر با

پس از شبیه‌سازی تراز سطح ایستابی با استفاده از سه مدل الگوریتم جنگل تصادفی، درخت تصمیم و شبکه عصبی مصنوعی، نتایج حاصل از آن‌ها با استفاده از معیارهای ارزیابی R2، RMSE و NS با تراز سطح ایستابی مشاهداتی آبخوان براساس روابط ۴-۶ مقایسه شد:

$$R^2 = \left(\frac{\sum_{i=1}^n (O_i - \bar{O})(P_i - \bar{P})}{\sqrt{\sum_{i=1}^n (O_i - \bar{O})^2} \sqrt{\sum_{i=1}^n (P_i - \bar{P})^2}} \right)^2 \quad (4)$$

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^N (P_i - O_i)^2}{N}} \quad (5)$$

$$\text{NS} = 1 - \frac{\sum_{i=1}^N (O_i - P_i)^2}{\sum_{i=1}^N (O_i - \bar{O})^2} \quad (6)$$

در روابط یادشده O_i ، P_i ، \bar{O} ، \bar{P} و N به ترتیب مقادیر تراز سطح ایستابی به صورت مشاهده‌ای و شبیه‌سازی شده توسط مدل، میانگین مقادیر مشاهده‌ای، میانگین مقادیر شبیه‌سازی شده و تعداد گام‌های زمانی را نشان می‌دهد. همچنین، در انتها دقت دسته‌بندی مدل جنگل تصادفی با استفاده از ماتریس اغتشاش^۱ و براساس رابطه ۷ که بیان‌کننده معیار دقت^۲ است، مورد ارزیابی قرار گرفت.

3. True Positive
4. False Negative
5. True Negative
6. False Positive

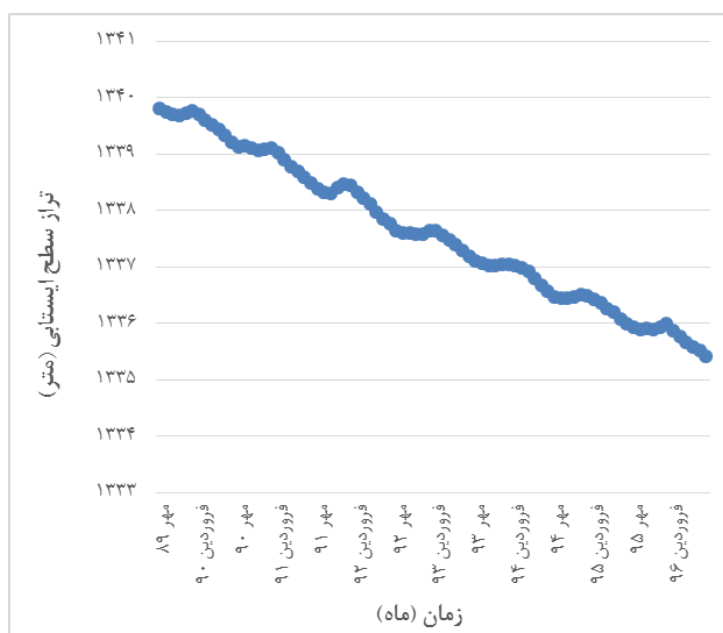
1. Confusion Matrix
2. Accuracy

شرایط با تعیین شیب خط روند، حذف روند براساس رابطه ۲ انجام شد. جدول ۲ نتایج بررسی وجود روند در داده‌های ورودی را نشان می‌دهد. شکل‌های ۴ و ۵ سری زمانی میانگین تراز سطح ایستابی چاه‌های مشاهده‌ای را در شرایط قبل و پس از حذف روند نشان می‌دهند.

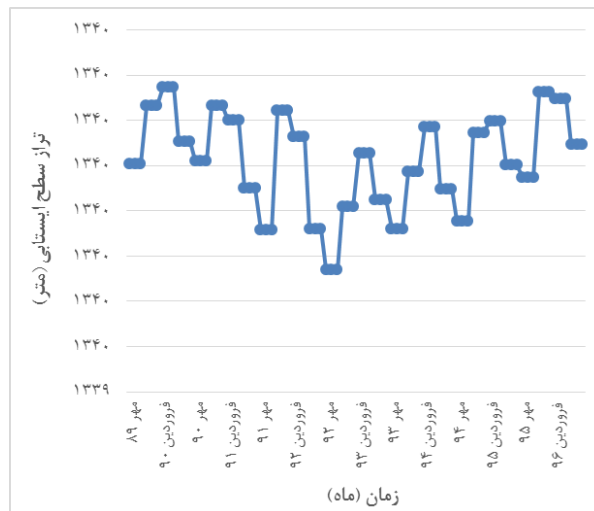
اطمینان ۹۵ درصد فرض صفر که در واقع نبود روند در سری زمانی و تصادفی بودن داده‌ها را نشان می‌دهد، قابل قبول است. همچنین، با انجام این آزمون روی پارامتر تراز سطح ایستابی در چاه مشاهده‌ای وجود روند به علت رد شدن فرض صفر ($p\text{-value} < 0.05$) اثبات شد که در این

جدول ۲. مقادیر p-value و شیب خط در آزمون من-کندال فصلی

شیب خط روند	p-value	پارامتر
۰/۰۹۲	۰/۱۶۱	دما
۰	۰/۴۸۴	بارندگی
-۰/۲۹۷	۰/۰۴۸	رطوبت
۰/۷۱۸	۰/۲۴۶	تبخیر
-۰/۰۹۷	< ۰/۰۰۰۱	PIZ 1
-۰/۱۱۶	< ۰/۰۰۰۱	PIZ 2
-۰/۱۵	< ۰/۰۰۰۱	PIZ 3
-۰/۱۰۱	< ۰/۰۰۰۱	PIZ 4
-۰/۰۸۴	< ۰/۰۰۰۱	PIZ 5
-۰/۱۲	< ۰/۰۰۰۱	PIZ 6
-۰/۲۱۵	< ۰/۰۰۰۱	PIZ 7
-۰/۱۳۶	< ۰/۰۰۰۱	PIZ 8
-۰/۱۱۶	< ۰/۰۰۰۱	PIZ 9
-۰/۴۷۶	< ۰/۰۰۰۱	PIZ 10
-۰/۱۴۹	< ۰/۰۰۰۱	PIZ 11



شکل ۴. سری زمانی میانگین تراز سطح ایستابی در چاه‌های مشاهده‌ای قبل از حذف روند



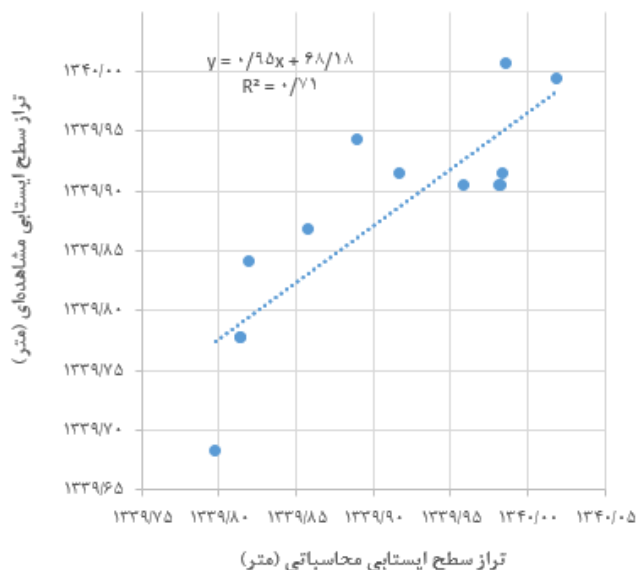
شکل ۵. سری زمانی میانگین تراز سطح ایستابی در چاه‌های مشاهده‌ای پس از حذف روند

پیش‌بینی سطح آب زیرزمینی دشت بوکان سازگار است [۱۸]. شکل ۷ نمودار پراکنش مقادیر تراز سطح ایستابی در دو حالت مشاهداتی و محاسباتی را نشان می‌دهد. از بررسی نمودار اهمیت پارامترهای ورودی در مدل‌سازی تراز آبخوان (شکل ۸) می‌توان دریافت که تراز سطح ایستابی چاه‌های مشاهده‌ای واقع در نواحی مرکزی دشت بیرجند نسبت به سایر چاه‌های مشاهده‌ای اثر بیشتری بر شبیه‌سازی الگوریتم جنگل تصادفی داشته است. همچنین، پارامترهای هواشناسی بارندگی، رطوبت، تبخیر و دما به‌ترتیب بیشترین اهمیت را در شبیه‌سازی تراز سطح ایستابی دارند. نتایج بررسی دقت دسته‌بندی الگوریتم جنگل تصادفی براساس ماتریس اغتشاش با مقدار ۷۳/۳۳ درصد بر دقت این دسته‌بندی تأکید دارد.

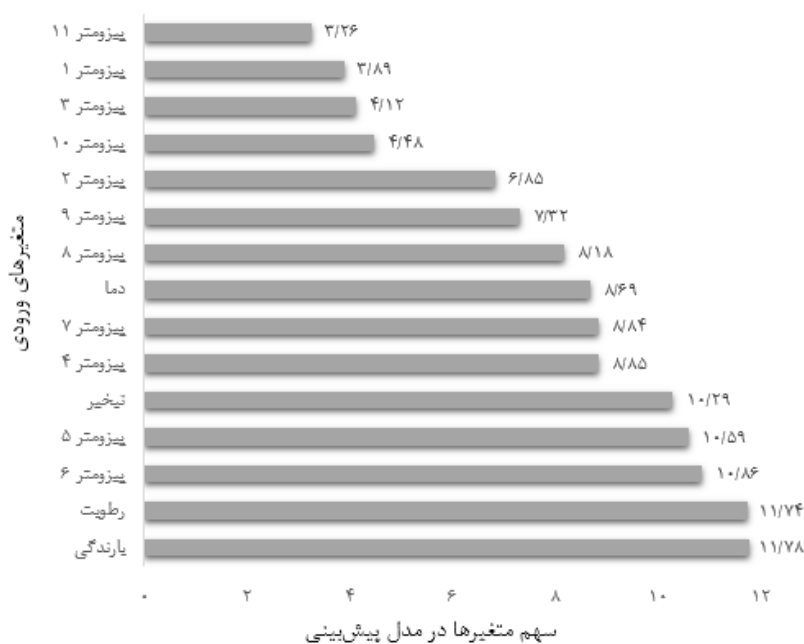
پس از حذف روند، تراز سطح ایستابی آبخوان با استفاده از مدل‌های جنگل تصادفی، درخت تصمیم و شبکه عصبی مصنوعی برآورد شد. بررسی نتایج به‌دست‌آمده از مقایسه تراز سطح ایستابی شبیه‌سازی‌شده با استفاده از روش الگوریتم جنگل تصادفی با تراز واقعی آبخوان نسبت به زمان (شکل ۶) نشان می‌دهد که این روش می‌تواند با دقت زیاد تراز سطح ایستابی آبخوان را برآورد کند و روند نوسانات تراز آبخوان را مدل کند. همچنین، مقادیر به‌دست‌آمده از معیارهای ارزیابی R^2 ، RMSE و NS که به‌ترتیب معادل ۰/۷۱۴، ۰/۰۰۳ و ۰/۵۹۸ است نیز بر این امر دلالت دارد. همچنین، نتایج حاصل از این تحقیق با نتایج پژوهش نوروزی و ندیری (۲۰۱۸) مبنی بر دقت زیاد روش جنگل تصادفی در



شکل ۶. مقایسه تراز ایستابی مشاهده‌ای و شبیه‌سازی شده با استفاده از الگوریتم جنگل تصادفی



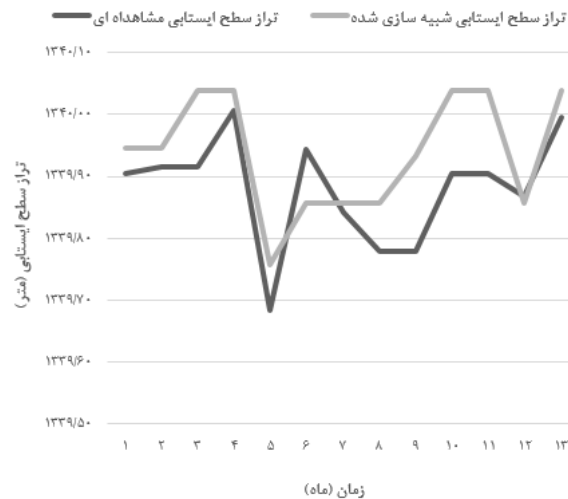
شکل ۷. نمودار پراکنش مقادیر تراز سطح ایستابی در دو حالت مشاهداتی و محاسباتی (جنگل تصادفی)



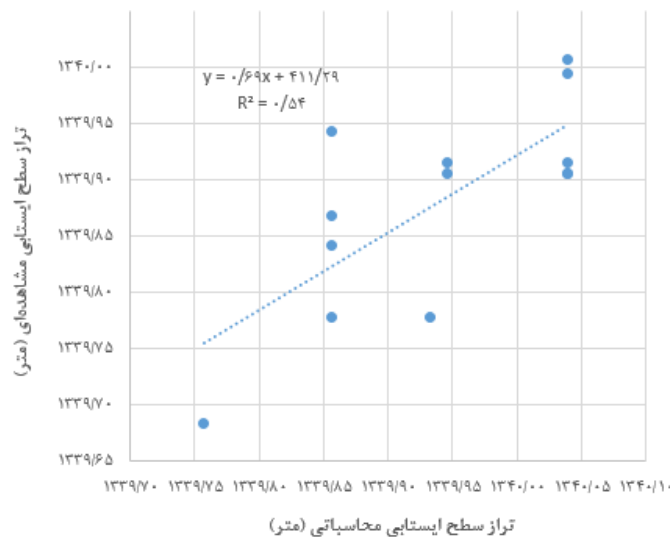
شکل ۸. نمودار اهمیت پارامترهای ورودی در مدل‌سازی تراز آبخوان

نمودار پراکنش مقادیر تراز سطح ایستابی در دو حالت مشاهداتی و محاسباتی با استفاده از مدل درخت تصمیم است که با توجه به پراکندگی زیاد نقاط در آن، دقت کم این مدل در شبیه‌سازی تراز آبخوان دشت بیرجند تأیید می‌شود. در مدل شبکه عصبی مصنوعی چاه‌های مشاهده‌ای ۷، ۸ و ۵ به ترتیب بیشترین تأثیر را بر نتایج مدل می‌گذارند.

همچنین، نتایج شبیه‌سازی با استفاده از روش درخت تصمیم با معیارهای ارزیابی معادل $R^2=0/5409$ ، $RMSE=0/072$ بر حسب متر و $NS=-0/187$ نیز عملکرد متوسط این روش را نسبت به روش جنگل تصادفی نشان می‌دهد. شکل ۹ تراز ایستابی مشاهده‌ای و شبیه‌سازی شده با استفاده از مدل درخت تصمیم را نسبت به زمان نشان می‌دهد. همچنین، شکل ۱۰ نشان‌دهنده



شکل ۹. مقایسه تراز ایستابی مشاهده‌ای و شبیه‌سازی شده با استفاده از مدل درخت تصمیم



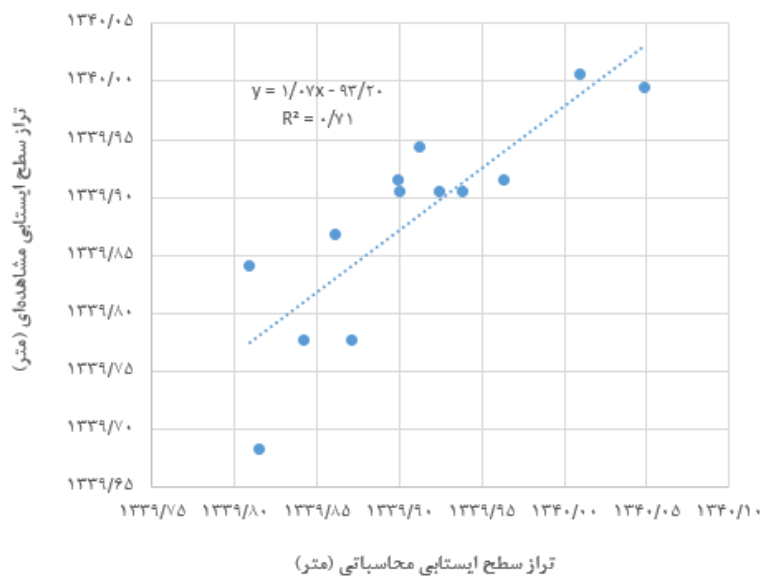
شکل ۱۰. نمودار پراکنش مقادیر تراز سطح ایستابی در دو حالت مشاهده‌ای و محاسباتی (درخت تصمیم)

پارامترها در پیش‌بینی تراز سطح ایستابی حائز اهمیت است. جدول ۳ نتایج معیارهای ارزیابی به‌دست‌آمده از مقایسه مدل‌های شبیه‌سازی یادشده با تراز مشاهده‌ای سطح ایستابی در دو بخش آموزش و آزمون را نشان می‌دهد. براساس نتایج حاصل، دو مدل الگوریتم جنگل تصادفی و شبکه عصبی مصنوعی نزدیک‌ترین جواب را نسبت به تراز واقعی آبخوان آزاد دشت بیرجند در مقایسه با روش درخت تصمیم ارائه کرده‌اند که این امر توانایی این دو مدل را برای شبیه‌سازی نوسانات تراز سطح ایستابی این آبخوان اثبات می‌کند.

بررسی نتایج تراز سطح ایستابی حاصل از مدل شبکه عصبی مصنوعی با معیارهای ارزیابی $R^2=0.7055$ ، $RMSE=0.003$ بر حسب متر و $NS=0.6046$ بر عملکرد سطح بالای این مدل در شبیه‌سازی تراز آبخوان دشت بیرجند و نزدیکی آن به نتایج مدل جنگل تصادفی اشاره دارد. همچنین، مقایسه دو مدل درخت تصمیم و شبکه عصبی مصنوعی نشان می‌دهد نتایج مدل شبکه عصبی تطابق بیشتری با تراز واقعی آبخوان دارد که این امر در شکل‌های ۱۱ و ۱۲ مشهود است. در مدل درخت تصمیم پارامترهای دما و بارندگی و نیز دو چاه مشاهده‌ای ۶ و ۴ نسبت به سایر



شکل ۱۱. مقایسه تراز ایستابی مشاهده‌ای و شبیه‌سازی شده با استفاده از مدل شبکه عصبی مصنوعی



شکل ۱۲. نمودار پراکنش مقادیر تراز سطح ایستابی در دو حالت مشاهداتی و محاسباتی (شبکه عصبی مصنوعی)

جدول ۳. معیارهای ارزیابی نتایج پیش‌بینی تراز سطح ایستابی

NS	RMSE (m)	R ²	مرحله	مدل شبیه‌سازی
۰/۵۹۸	۰/۰۰۳	۰/۷۱۴	آزمون	الگوریتم جنگل تصادفی
۰/۸۹۹	۰/۰۰۱۲	۰/۹۰۹	آموزش	
-۰/۰۱۸۷	۰/۰۰۷۲	۰/۵۴۰۹	آزمون	درخت تصمیم
۰/۷۷۶۸	۰/۰۰۲۸	۰/۷۷۶۸	آموزش	
۰/۶۰۴۶	۰/۰۰۳	۰/۷۰۵۵	آزمون	شبیه عصبی مصنوعی
۰/۸۳۷۸	۰/۰۰۲	۰/۸۳۷۸	آموزش	

نتیجه‌گیری

امروزه، کمبود بارش و وجود خشکسالی‌های پی‌درپی، رشد روزافزون جمعیت و افزایش تقاضا در بخش‌های مختلف شرب، صنعت و کشاورزی به‌ویژه در مناطق خشک و نیمه‌خشک سبب برداشت‌های بی‌رویه از منابع آب زیرزمینی و افت تراز آبخوان‌ها شده است. از جمله روش‌های حفاظت از این منابع آبی، بررسی و تعیین سطح ایستابی آبخوان‌ها در دوره‌های مختلف و ارائه راهکارهای مدیریتی برای مقابله با کاهش تراز، نشست دشت‌ها و مشکلات ناشی از آن است. تا کنون در راستای پیش‌بینی و تعیین تراز سطح ایستابی مدل‌های مختلفی مطرح شده است که می‌تواند کمک شایانی در مدیریت آبخوان‌ها ارائه دهد. در این مطالعه کارایی الگوریتم جنگل تصادفی در شبیه‌سازی تراز سطح ایستابی آبخوان دشت بیرجند با در نظر گرفتن پارامترهای تراز سطح ایستابی ۱۱ پیژومتر (چاه‌های مشاهده‌ای)، دما، بارندگی، رطوبت و تبخیر به صورت ماهانه طی سال‌های آبی ۱۳۸۹-۱۳۹۰ تا ۱۳۹۵-۱۳۹۶ در نرم‌افزار آماری R بررسی شد و در انتها، نتایج به‌دست‌آمده از آن با نتایج دو روش مدل درخت تصمیم و شبکه عصبی مصنوعی (خروجی نرم‌افزار آماری R) مقایسه شد. نتایج معیارهای ارزیابی نشان داد تراز سطح ایستابی شبیه‌سازی‌شده با استفاده از الگوریتم جنگل تصادفی با $NS=0/598$ ، $RMSE=0/03$ ، $R^2=0/714$ بر حسب متر و $NS=0/6046$ ، $RMSE=0/03$ بر حسب متر و $R^2=0/7055$ مورد مدل شبکه عصبی مصنوعی نیز با $NS=0/6046$ ، $RMSE=0/03$ بر حسب متر و $R^2=0/7055$ مورد استفاده در شبیه‌سازی نشان داد نتایج این الگوریتم بیشتر از اطلاعات ورودی چاه‌های مشاهده‌ای واقع در نواحی مرکزی آبخوان و نیز پارامترهای هواشناسی تأثیر می‌پذیرد. بنابراین، از مقایسه نتایج الگوریتم جنگل تصادفی با دو مدل درخت تصمیم و شبکه عصبی مصنوعی می‌توان دریافت که دو مدل الگوریتم جنگل تصادفی و شبکه عصبی مصنوعی نسبت به مدل درخت تصمیم عملکرد سطح بالاتری را نشان می‌دهند و از قابلیت استفاده در آبخوان مورد نظر را برخوردارند.

منابع

- [1]. Rajae T, Zeynivand A. Modeling of groundwater level using ANN-Wavelet Hybrid model (Case Study: Sharif Abad Plain). Journal of Civil and Environmental Engineering. 2015; 44(4): 51-63. [Persian]
- [2]. Mirmorsley N, Karbasi M. Comparison of the J48, Random Forest and Tree Random algorithms efficiency in predicting bed form in sandy rivers. Second Iranian National Hydrology Conference. 2017; Shahrekord University, Shahrekord, Iran. [Persian]
- [3]. Rajae T, Mirbagheri S.A. Suspended sediment model in rivers using artificial neural networks. J. Civil Engin. 2009; 21(1): 27-43. [Persian]
- [4]. Rajae T, Ebrahimi H. Application of wavelet-neural network model for forecasting groundwater level time series with non-stationary and nonlinear characteristics. J. of Water and Soil Conservation. 2016; 22(5): 99-115. [Persian]
- [5]. Altunkaynak A. Forecasting surface water level fluctuations of Lake Van by artificial neural networks. Water Resour. Manage. 2007; 21 (2): 399-408.
- [6]. Li Y, Zhang Q, Yao J, Werner A.D, Li X. Hydrodynamic and hydrological modeling of the Poyang Lake catchment system in China. J. Hydrol. Eng. 2013; 19 (3): 607-616.
- [7]. Li B, Yang G, Wan R, Dai X, Zhang Y. Comparison of random forests and other statistical methods for the prediction of lake water level : a case study of the Poyang Lake in China. Hydrology Research. 2016; 69-83.
- [8]. Khalili Naft Chali A, Shahidi A, khashei siuki A. Comparison of Lazy Algorithms and M5 model to estimate groundwater level (Case Study: Plain Neyshabur). JWSS. 2017; 21 (3): 15-26. [Persian]
- [9]. Mirhashemi S.H, Haghghat jou P, Mirzaei F, Panahi M. Using CART algorithm in predicting groundwater table fluctuations inside and outside of an irrigation system (case study: irrigating area of Qazvin). Iranian Journal of Soil and Water Research. 2018; 49(2): 385-395. [Persian]
- [10]. Mohtasham M, Dehghani A.A, Akbarpour A, Meftah M, Etebari B. Oundwater level determination by using Artificial Neural Network (Case study: Birjand Aquifer). Iranian Journal of Irrigation and drainage. 2010; 4(1): 1-10. [Persian]

- [11]. Khashei-Siuki A, Ghahraman B, Kouchakzadeh M. Comparison of ANN, ANFIS and Regression models to estimate groundwater level of Neyshaboor Aquifer. *Iranian Journal of Irrigation and Drainage*. 2013; 1(7): 10-22. [Persian]
- [12]. Saeedi Razavi B, Arab A. Groundwater Level Prediction of Ajabshir Plain using Fuzzy Logic, Neural Network Models and Time Series. *Hydrogeology*. 2019; 3(2): 69-81. [Persian]
- [13]. Jabalbarez B, Malekian A. Comparison of the performance of artificial neural networks and gene expression to predict the groundwater level in arid and semi-arid areas (Case study: Jiroft plain). *Iranian Journal of Range and Desert Research*. 2019; 26(2): 292-301. [Persian]
- [14]. Nayak P.C, Satyaji Rao Y.R, Sudheer K.P. Groundwater level forecasting in a shallow aquifer using artificial neural network approach. *Water Resources Management*. 2006; 2(1): 77-99.
- [15]. Srekanth P.D, Geethanjali N, Sreedevi P.D, Ahmed S, Ravi Kumar N, Kamala Jayanthi P.D. Forecasting groundwater level using artificial neural networks. *Current Science*. 2009; 96: 933- 939.
- [16]. Sun Y, Wendi D, Kim D.E, Liong S.Y. Technical note: Application of artificial neural networks in groundwater table forecasting - a case study in a Singapore swamp forest. *Hydrology and Earth System Sciences*. 2016; 20(4): 1405-1412.
- [17]. Noruzi H, Nadiri A.A, Asgharimoghaddam A, Gharekhani M. Prediction of Transmissivity of Malikan Plain Aquifer Using Random Forest Method. *Water and Soil Science*. 2017; 27(2): 61-75. [Persian]
- [18]. Norouzi H, Nadiri A. Groundwater Level Prediction of Boukan Plain using Fuzzy Logic, Random Forest and Neural Network Models. *Journal of Range & Watershed Management*. 2018; 71(3): 829-845. [Persian]
- [19]. Wang X, Liu T, Zheng X, Peng H, Xin J, Zhang B. Short-term prediction of groundwater level using improved random forest regression with a combination of random features. *Applied Water Science*. 2018; 8(5): 1-12.
- [20]. Hamraz B.S, Akbarpour A, Pourreza Bilondi M. Assessment of parameter uncertainty of MODFLOW model using GLUE method (Case study: Birjand plain). *Journal of Water and Soil Conservation*. 2016; 22(6): 61-79. [Persian]
- [21]. Farpoor A, Ramezani Y, Akbarpour A. Numerical Simulation of Chromium Changes Trend in Aquifer of Birjand Plain. *Iranian Journal of Irrigation and Drainage*. 2018; 12(5): 1203-1216. [Persian]
- [22]. Breiman L. Random forests. *Mach Learn*. 2001; 45(1):5-32.
- [23]. Kotsiantis S, Pintelas P. Combining bagging and boosting. *Journal of Computational Intelligence*. 2004; 1(4): 324-333.
- [24]. Guo L, Chehata N, Mallet C, Boukir S. Relevance of airborne lidar and multispectral image data for urban scene classification using Random Forests. *ISPRS J Photogramm Remote Sens*. 2011; 66(1): 56-66.
- [25]. Rodriguez-Galiano V, Mendes M.P, Garcia-Soldado M.J, Chica-Olmo M, Ribeiro L. Predictive modeling of groundwater nitrate pollution using Random Forest and multisource variables related to intrinsic and specific vulnerability: A case study in an agricultural setting (Southern Spain). *Science of the Total Environment*. 2014; 189-206.
- [26]. Talebi A, Akbari Z. Investigation of ability of decision Trees model to estimate river suspended sediment (Case Study: Ilam Dam Basin). *J. Sci. & Technol. Agric. & Natur. Resour. Water and Soil Sci*. 2013; 17(63): 109-121. [Persian]
- [27]. Fallahi M.R, Varvani H, Golian S. Precipitation prediction using tree regression model to flood control. *Fifth National Conference on Watershed Management and Soil and Water Resources Management*. 2012; Kerman, Iran. [Persian]
- [28]. Ghafari G.A, Vafakhah M. Simulation of rainfall-runoff process using Artificial Neural Network and adaptive Neuro-Fuzzy Interface System (Case Study: Hajighoshan Watershed). *Journal of Watershed Management Research*. 2013; 4(8): 120-136. [Persian]
- [29]. Yue, S, Wang C. Y. The Mann-Kendall test modified by effective sample size to detect trend in serially correlated hydrological series, *Water Res.Manage*. 2004; 18: 201-218.
- [30]. Ghodoosi1 M, Morid S, Delavar M. Comparison of detrending methods for the temperature and precipitations time series. *Journal of Agricultural Meteorology*. 2013; 1(2): 32-45. [Persian]
- [31]. Ahmadi F, Radmanesh F. Trend Analysis of Monthly and Annual Mean Temperature of the Northern Half of Iran Over the Last 50 Years. *Journal of Water and Soil*. 2014; 28(4): 855-865. [Persian]