

مقاله علمی-پژوهشی

ارزیابی توانایی مدل‌های درختی جنگل تصادفی و M5P در برآورد سطح ایستابی دشت سراوان با استفاده از داده‌های هواشناسی

سیاوش سپاهیان^{۱*}، سید آرمان هاشمی منفرد^۲، عباسعلی قادری^۳

تاریخ دریافت: ۱۴۰۲/۰۷/۲۳ تاریخ پذیرش: ۱۴۰۲/۱۰/۰۸

چکیده

آب‌های زیرزمینی یکی از کلیدی‌ترین منابع آب در سطح کره خاکی هستند که در بخش کشاورزی، صنعت و شرب دارای اهمیت بوده و ضامن حیات برای نسل آینده به شمار می‌آیند. در این تحقیق از ۲۲۷ سری داده مربوط پارامترهای هواشناسی و آب زیرزمینی دشت سراوان در استان سیستان و بلوچستان جهت پیش‌بینی و برآورد سطح ایستابی آب زیرزمینی با استفاده مدل‌های هوشمند یادگیری ماشین استفاده شد. از این تعداد سری داده، ۱۵۹ سری برای استفاده در بخش آموزش مدل‌ها و ۶۸ سری برای آزمون مدل‌ها در سناریوهای طراحی شده تخصیص یافت. عملکرد مدل‌های جنگل تصادفی و M5P در برآورد متوسط سطح ایستابی دشت سراوان مورد ارزیابی قرار گرفت. عملکرد مدل‌ها با استفاده از شاخص‌های آماری ضریب همبستگی (R)، خطای مجذور میانگین مربعات (RMSE) و خطای بایاس میانگین (MBE) مقایسه شد. نتایج به دست آمده نشان داد هر دو مدل به کاررفته از توانایی خوبی در مدل‌سازی سطح ایستابی برخوردارند، اما مدل M5P با ضریب همبستگی بالاتر خطاهای کمتری نسبت به مدل جنگل تصادفی دارد. سناریوی چهارم مدل M5P با مقادیر ضریب همبستگی، MBE و RMSE به ترتیب ۰/۹۶۱، ۰/۰۰۰۱- و ۰/۷۲۷ در بخش آموزش و ۰/۸۷۱، صفر و ۰/۲۶۵ در بخش آزمون به‌عنوان بهترین مدل در بین سناریوهای دو مدل انتخاب شد. پیشنهاد می‌شود سطح ایستابی دشت سراوان با سایر روش‌های هوش مصنوعی و یادگیری ماشین نیز ارزیابی و با مطالعه حاضر مقایسه شود.

واژه‌های کلیدی: آبخوان، تراز آب زیرزمینی، درخت تصمیم، جنگل تصادفی، مدیریت منابع آب

مقدمه

دقیق از سطح ایستابی دشت می‌باشد که از طریق مدل‌های غیرخطی انجام می‌پذیرد. استفاده از مدل‌های یادگیری ماشین و هوش مصنوعی در پیش‌بینی منابع آبی از جمله آب‌های زیرزمینی رشد چشمگیری در مهر و موم‌های گذشته داشته است. ابراهیمی و همکاران (۱۳۹۷) سطح آب زیرزمینی دشت بیرجند را با استفاده از مدل شبکه‌های بیزین پیش‌بینی کردند. آن‌ها با بررسی ۱۰ رویکرد تحلیل حساسیت برای پیش‌بینی سطح ایستابی آبخوان، مشاهده کردند که پارامتر سطح آب زیرزمینی ماه فعلی اثرگذاری بالایی در افزایش دقت مدل دارد.

ندیری و همکاران (۱۳۹۷) با استفاده از ۵ پارامتر ورودی بارش، دما، تبخیر، دبی خروجی و تراز آب زیرزمینی در یک زمان قبل به مدل‌سازی تراز آب زیرزمینی دشت تبریز با استفاده از مدل ترکیبی SOM-ANN پرداختند. بر اساس تغییرات تراز ۱۴ ساله داده به چهار کلاس دسته‌بندی و با استفاده از شبکه عصبی مصنوعی پیش‌بینی انجام شد و میانگین خطا برای هر چهار کلاس مورد ارزیابی قرار گرفت. آن‌ها بیان کردند با توجه به وضعیت پیچیده زیرسطحی در

استفاده بی‌رویه از آب‌های زیرزمینی در دهه‌های اخیر سبب اختلال در تعادل طبیعی آن شده و سطح ایستابی در بسیاری از آبخوان‌های کشور افت کرده است. استفاده نامتعارف آب‌های زیرزمینی به دلایلی نظیر استفاده از روش‌های ناصحیح و عدم مدیریت آبیاری و حفر چاه‌های عمیق مربوط است که می‌تواند سبب تخریب سرزمین شود (صادق‌روش و همکاران، ۲۰۱۵). آگاهی از وضعیت آب زیرزمینی و مدیریت منابع آبی بهینه مستلزم پیش‌بینی

۱- دانشجوی دکتری، گروه مهندسی عمران، دانشکده مهندسی شهید نیکبخت،

دانشگاه سیستان و بلوچستان، ایران

۲- دانشیار گروه مهندسی عمران، دانشکده مهندسی شهید نیکبخت، دانشگاه

سیستان و بلوچستان، ایران

۳- استادیار گروه مهندسی عمران، دانشکده مهندسی شهید نیکبخت، دانشگاه

سیستان و بلوچستان، ایران

(*- نویسنده مسئول: (Email: s.sepaheyan@gmail.com)

کردند (Stanley Raj, 2017). برزگر و همکاران (۲۰۱۷) با استفاده از مدل‌های ترکیبی موجک و شبکه عصبی، پیش‌بینی نوسانات سطح آب زیرزمینی را انجام دادند. استفاده از شبکه‌های تقویت عصبی WA عملکرد بهتری برای پیش‌بینی سطح آب زیرزمینی در مقایسه با مدل‌های مبتنی بر شبکه عصبی داشته‌اند.

ناصری (۲۰۱۸) افت سطح آب زیرزمینی را به مدت ۴۵ سال با روش الگوبندی، به‌منظور تحلیل تغییرات سطح آب زیرزمینی در یک دشت ارزیابی کرد و الگوی اتورگرسیو تلفیقی میانگین متحرک مناسب‌ترین الگو تشخیص داده شد.

رادمنش و همکاران با استفاده از داده‌های ماهانه پیرومتری آبخوان دشت لور- اندیمشک، به مدت ۱ سال، و محاسبات روی مدل مفهومی آب زیرزمینی MODFLOW، برنامه‌ریزی بیان ژن و سیستم استنتاج فازی عصبی تطبیقی برای مدل‌سازی نوسانات سطح ایستابی آبخوان اجرا کردند. نتایج نشان داد که سرعت اجرای مدل برنامه‌ریزی بیان ژن سریع‌تر از مدل سیستم استنتاج فازی عصبی تطبیقی بوده و هر دو مدل از نظر کارایی زمانی بهتر از مدل مفهومی عمل کردند (Radmanesh et al., 2020).

بر اساس منابع بررسی‌شده سطح ایستابی دشت سراوان با استفاده از روش‌های هوش مصنوعی پیش‌بینی نشده است. همچنین روش جنگل تصادفی در این خصوص در کشور بکار نگرفته شده است. لزوم دسترسی به روش‌های زودیافت ایجاب می‌کند تا روش‌های هوش مصنوعی و یادگیری ماشین نوین و کمتر بکار گرفته شده مورد ارزیابی قرار گیرند. لذا بنابر ضرورت موجود در زمینه مدل‌سازی هوشمند موضوعات مدیریت آب، هدف از این مطالعه ارائه مدلی مبتنی بر هوش مصنوعی و یادگیری ماشین در راستای برآورد سطح ایستابی دشت سراوان است.

داده‌ها و روش‌ها

این پژوهش در حوضه آبریز سراوان استان سیستان و بلوچستان در جنوب شرق کشور انجام شده است. در شکل ۱ موقعیت حوضه آبریز سراوان در استان سیستان و بلوچستان نشان داده شده است. دشت سراوان یک حوضه کاملاً کشیده و کم عرض است که زهکشی اصلی آن در طول حوضه از حاشیه جنوبی دشت، از شمال غرب به جنوب شرق به سمت سراوان ادامه دارد. این حوضه از ۲۴ کیلومتری غرب گشت با رودخانه پسیل آغاز و با گذشتن از رودخانه‌های سیمیش، مک کورو روتک به رودخانه ماشکید پیوسته و نهایتاً به سمت پاکستان هدایت می‌شود. متوسط ارتفاع از سطح دریا در این منطقه ۱۱۹۵ متر از سطح دریا و متوسط بارندگی سالیانه بلندمدت آن حدود ۳۵ میلی‌متر می‌باشد.

خصوص پیش‌بینی از طریق روش دسته‌بندی هوشمند غیر نظارتی می‌تواند روش مناسبی برای تفکیک منطقه مطالعاتی به چند قسمت همگن و در نتیجه کاهش مقادیر خطا شود.

علیمرزایی و همکاران (۱۳۹۸) به شبیه‌سازی سطح ایستابی آب زیرزمینی دشت سرخس با استفاده از داده‌های سطح ایستابی ۱۸ حلقه چاه و بارش و تبخیر پتانسیل ماهانه از طریق ترکیب روش‌های هوشمند با زمین‌آمار پرداختند. در این پژوهش ترکیب مدل MLP با درون‌یابی کریجینگ به‌عنوان راه‌حل مناسب و با هزینه کم به‌منظور شبیه‌سازی تراز آب زیرزمینی دشت سرخس ارزیابی شد.

حسینی صومعه (۱۳۹۹) در پژوهشی تغییرات سطح آب زیرزمینی بر اساس روش‌های مبتنی بر هوش مصنوعی را در دشت زاوه تربیت حیدریه مدل‌سازی کردند. آن‌ها از RMSE برای ارزیابی مدل‌ها استفاده کردند.

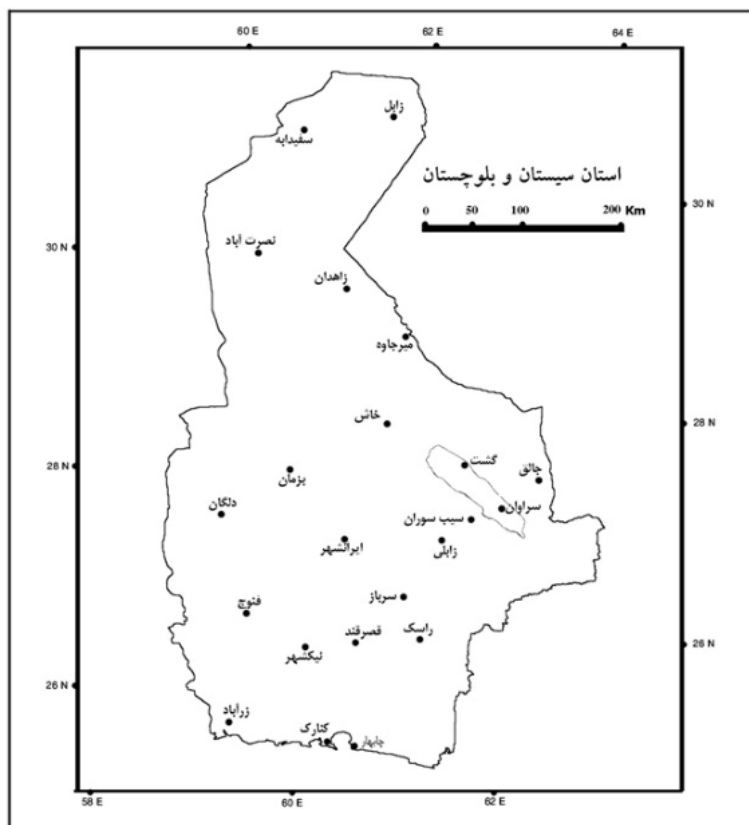
پور صالحی و همکاران (۱۴۰۰) عملکرد الگوریتم جنگل تصادفی در پیش‌بینی نوسانات سطح ایستابی در مقایسه با دو مدل درخت تصمیم و شبکه عصبی مصنوعی را در آبخوان آزاد دشت بیرجند مورد مطالعه و بررسی قرار دادند. نتایج حاصل از این تحقیق نشان داد الگوریتم جنگل تصادفی توانایی نسبتاً خوبی در شبیه‌سازی تراز سطح ایستابی آبخوان دارد. مقایسه الگوریتم جنگل تصادفی با دو مدل درخت تصمیم و شبکه عصبی مصنوعی حاکی از برتری آن نسبت به مدل درخت تصمیم و هم‌راستا بودن نتایج آن با شبکه عصبی مصنوعی می‌باشد. همچنین آن‌ها بیان کردند که پارامترهای هواشناسی بارندگی و رطوبت در شبیه‌سازی تراز سطح ایستابی آبخوان نقش موثرتری نسبت به سایر پارامترهای هواشناسی داشتند.

دالیاکوپولوس و همکاران با استفاده از شبکه‌های عصبی مصنوعی با معماری پیشرو و الگوریتم آموزش لونیگ-مارکوات، مقادیر سطح آب زیرزمینی دره مسارا در جزیره کرت در جنوب یونان را تا ۱۸ ماه آینده پیش‌بینی کردند (Daliakopoulos et al., 2005).

ندیری و همکاران در پژوهشی از مدل‌های هوشمند منطق فازی تاکاگی-سوگنو، شبکه عصبی مصنوعی و ANFIS برای تخمین هدایت هیدرولیکی آبخوان تسوج-آذربایجان شرفی استفاده کردند (Nadiri et al., 2014).

سان و همکاران به‌منظور پیش‌بینی سطح آب زیرزمینی جنگل‌های باتلاقی سنگاپور با کمک شبکه عصبی از معماری پرسپترون چندلایه با الگوریتم لونیگ-مارکوات و فقط مقادیر سطح آب برکه‌ها و میزان بارندگی استفاده کردند. نتایج حاکی از این است که صحت مقادیر پیش‌بینی شده تا ۷ روز آینده بود ولی با طولانی‌تر شدن زمان پیش‌بینی دقت مقادیر کاهش یافت (Sun et al., 2015).

استتلی راج از شبکه عصبی مصنوعی برای پیش‌بینی تراز آب زیرزمینی در آبخوان کانیاکومری، تامیل نادو، هند استفاده



شکل ۱- موقعیت حوزه آبخیز سراوان در استان سیستان و بلوچستان

از طریق جایگزینی تولید می‌کند. این روش هم در زمینه طبقه‌بندی و هم در زمینه رگرسیون کاربرد دارد. با توجه به ویژگی تصادفی و نا همبسته الگوریتم جنگل تصادفی، این روش قادر به برقراری ارتباط بین پارامترهای ورودی و خروجی است، حتی زمانی که رابطه آن‌ها پیچیده و غیرخطی باشد (He et al., 2016; Hong et al., 2016). جنگل تصادفی یک روش گروهی است که چند الگوریتم درختی را برای تولید یک پیشبینی مکرر از هر پدیده ترکیب می‌کند. به‌طور کلی، درخت تصمیم منفرد مستعد برازش اضافی بوده و قدرت تعمیم‌پذیری کمی دارد. اصول کلی فن‌های آموزش گروهی بر اساس این فرضیه استوار است که دقت آن‌ها از سایر الگوریتم‌های آموزشی بالاتر است، چون ترکیبی از چند مدل پیشبینی دقیق‌تر از یک مدل است و گروه‌ها باعث کاهش نقاط ضعف طبقه‌ها می‌شوند (Kotsiantis and Pintelas, 2004; Pourghasemi and Kerle, 2016; Rahmati et al., 2016). در هنگام تشکیل یک درخت تصمیم، تغییرات کم در الگوهای یادگیری می‌تواند باعث تغییرات اساسی در ساختار آن درخت گردد. 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 به علت عدم وجود توزیع نرمال داده‌های استفاده‌شده، در

در این تحقیق از اطلاعات میانگین ماهانه هواشناسی و آب زیرزمینی دشت سراوان برای مدل‌های مورد ارزیابی بهره گرفته شد. اطلاعات مربوط به ورودی مدل شامل پارامترهای هواشناسی میانگین دمای حداکثر، میانگین دمای حداقل، دمای میانگین، میانگین رطوبت نسبی، میزان بارندگی، تبخیر ماهانه، میانگین سرعت باد و مجموع ساعت آفتابی و پارامتر آب زیرزمینی سطح ایستابی ماه قبل است که جهت پیشبینی سطح ایستابی در دشت سراوان بکار گرفته شده است. بر این اساس ۲۲۷ سری داده مربوط به مهر و موم‌های آبی ۸۲-۱۳۸۱ تا ۱۳۹۹-۱۴۰۰ استخراج و جهت پیشبینی سطح ایستابی دشت سراوان مورداستفاده قرار گرفت. از این تعداد داده ۱۵۹ سری داده برای بخش آموزش و ۶۸ سری داده برای بخش آزمون مدل‌ها اختصاص یافت.

جنگل تصادفی

الگوریتم جنگل تصادفی در سال ۲۰۰۱ توسط بریمن^۱ ارائه شد. این روش یک روش یادگیری ترکیبی است که چندین درخت تصمیم را با استفاده از یک زیرمجموعه به‌طور تصادفی انتخاب‌شده از نمونه‌ها

1 Breiman

گره‌ها و برگ‌هاست. در ریشه درخت تصمیم صفت قرار دارد و داده‌ها بر اساس آن صفت تفکیک می‌شوند. این تقسیم‌بندی تا جایی ادامه دارد که داده‌ها دیگر جداسازی نباشند، یا در فرایند رسیدن به هدف نیازی به تقسیم‌بندی بیشتر احساس نشود (Hand et al., 2011). در درخت تصمیم هر گره نشانده‌اند صفت مشخص و شاخه‌ها بازه‌ای از مقادیر معلوم برای صفت هستند. درختان تصمیم استقرایی بر اساس مجموعه‌ای از داده‌های معلوم، درخت تصمیم استنتاجی را تشکیل می‌دهند و این درخت به صورت مجموع قوانینی جهت پیشینی خصوصیات معلوم می‌تواند بکار گرفته شود (عشقی و همکاران، ۲۰۱۶). معیار تقسیم برای انتخاب بهترین متغیر برای جداسازی بخشی از مقادیر دسته‌ای که به یک گره خاص می‌رسند، طبق انحراف معیار مقادیر دسته‌ای و احتساب کاهش موردنظر در این خطا به عنوان نتیجه آزمون هر صفت در آن گره است. کاهش انحراف معیار بر اساس رابطه (۱) محاسبه می‌شود (Wang and Witten, 1997):

$$SDR = sd(T) - \sum \frac{|T_i|}{|T|} sd(T_i) \quad (1)$$

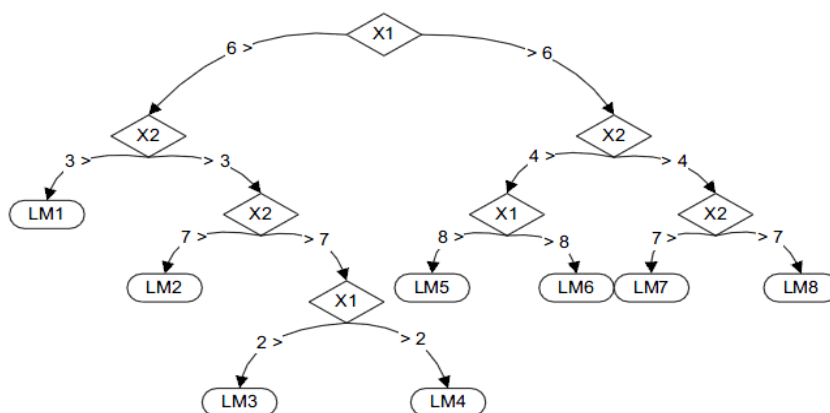
که در آن T دسته‌ای از نمونه‌ها هستند که به گره می‌رسند، T_i مجموعه‌هایی هستند که از تقسیم کردن گره بر اساس صفت انتخابی به دست آمده‌اند و sd انحراف از معیار است.

بعد از انجام کلیه آزمون‌های ممکن بر روی صفات، آنگاه M5P آن صفتی را که بتواند کاهش خطای بیشتری در گره ایجاد نماید را به عنوان صفت جداکننده و شاخه زنی انتخاب می‌نماید. فرآیند تقسیم داده‌ها تا وقتی ادامه خواهد داشت که تغییرات مقادیر متغیر هدف نمونه‌های رسیده به یک گره اندک شده و یا تعداد نمونه‌های کمی در گره باقی بماند (Quinlan, 1992). در شکل ۲ ساختار مدل M5P و در شکل ۳ تقسیم‌بندی فضای نمونه‌ها برای الگوریتم مذکور نشان داده شده است.

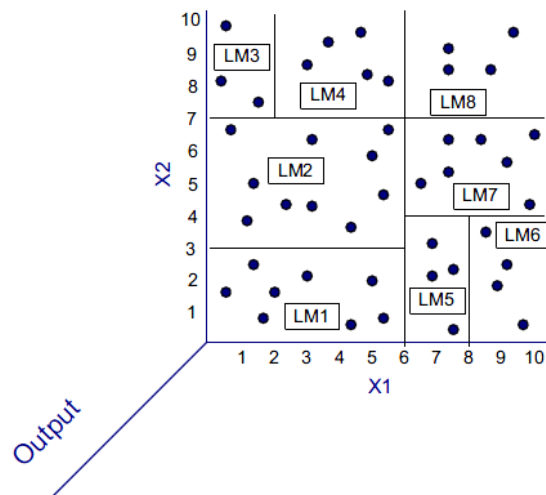
تجزیه و تحلیل گنجانده و ترکیب کند. ترکیب درخت‌های تصمیم را روش‌های دسته‌جمعی می‌گویند. یک مجموعه دسته‌جمعی شامل تعدادی یادگیرنده است که به آن‌ها یادگیرنده‌های پایه گفته می‌شود. یادگیرنده‌های پایه معمولاً از داده‌های آموزشی به وسیله یک الگوریتم یادگیری پایه ساخته می‌شوند که این الگوریتم یادگیری پایه می‌تواند درخت تصمیم، شبکه عصبی یا الگوریتم‌های یادگیری دیگر باشد. قابلیت تعمیم یک مجموعه، اغلب قوی‌تر از یادگیرنده‌های پایه است و در واقع روش‌های مجموعه‌ای بیشتر به دلیل داشتن توانایی در تقویت یادگیرنده‌های ضعیف، قابل قبول هستند (Schapire, 1990). در این الگوریتم از چندین درخت تصمیم استفاده می‌شود که به هر کدام از درخت‌ها، یک زیرمجموعه‌ای از اطلاعات داده می‌شود. این درخت‌ها با همین داده‌های زیرمجموعه، می‌توانند تصمیم بگیرند و مدل طبقه‌بندی خود را ایجاد کنند (Senagi et al., 2017).

مدل درختی M5P

کوین لان اولین بار این الگوریتم را در سال ۱۹۹۲ مطرح کرد. مدل درختی M5P یکی از رایج‌ترین الگوریتم‌های درخت تصمیم است که از جمله مدل‌های با کارایی مناسب و هزینه محاسباتی اندک برای شبیه‌سازی و مدل‌سازی دستگاه‌های پیچیده است. این مدل از روش‌های نسبتاً جدید و توانای هوش محاسباتی با قابلیت مناسب است که توسط محققان بسیاری مورد استفاده قرار گرفته است (Etemad-Shahidi; Bonakdar and Etemad-Shahidi, 2011; Rahimikhoob, Sattari et al., 2013; and Mahjoobi, 2009; 2014; Wolf and Willems, 2014). درخت‌های تصمیم فاصله میان تولید داده و فهم آن را کاهش داده و راهکارهایی برای کشف الگوی حاکم بر موضوع ارائه می‌نمایند. مهم‌ترین مزیت آن‌ها ایجاد قوانین قابل فهم است. از ساختارهای درختی جهت پیشگویی یا دسته‌بندی استفاده می‌شود که برای این منظور از یکسری قوانین سلسله‌ای بهره گرفته می‌شود. ساختار درختی متشکل از شاخه‌ها،



شکل ۲- ساختار سلسله مراتبی مدل درختی M5P



شکل ۳- تقسیم‌بندی فضای نمونه‌ها توسط الگوریتم مدل درختی MSP

اینکه از مدل‌های مختلف هوش مصنوعی و یادگیری ماشین بتوان در راستای مدل‌سازی و تخمین در زمینه‌های مختلف علوم آب و محیط‌زیست استفاده کرد بایستی ابتدا سناریوهای مناسب تهیه شود. در این تحقیق بر اساس آنکه از داده‌های هواشناسی به‌منظور شبیه‌سازی سطح ایستابی دشت سراوان استفاده شده است، بر این اساس ورودی‌های در نظر گرفته شده تحت ۵ سناریو طراحی شده تا برای مدل‌سازی استفاده شود. در جدول ۲ سناریوهای در نظر گرفته شده برای ارزیابی مدل‌سازی سطح ایستابی ارائه شده است.

نتایج مدل جنگل تصادفی

نتایج به‌دست‌آمده از مدل جنگل تصادفی نشان می‌دهد که سناریوهای در نظر گرفته شده عملکرد خوبی در بخش آموزش و آزمون دارد. مقادیر خطای MBE و RMSE در محدوده قابل‌قبول بود. همچنین مقادیر ضریب همبستگی در بخش آموزش و آزمون به‌طور کلی از سناریوی اول به سناریوی پنجم افزایشی است. بر همین اساس شاخصه‌ای آماری نشان‌دهنده برتری سناریوی پنجم در مدل جنگل تصادفی در بخش آموزش و آزمون هست، اما در عین حال سناریوی چهارم هم اختلاف چندانی با آن ندارد. در شکل ۴ پراکنش داده‌های مشاهداتی در مقابل داده‌های برآوردی در بخش آزمون سناریوی پنجم به‌عنوان سناریوی برتر نشان داده شده است.

نتایج مدل MSP

نتایج به‌دست‌آمده از مدل MSP نشان‌دهنده نزدیکی نتایج در سناریوهای مختلف ارائه شده است (جدول ۴). در این مدل مقادیر خطا در همه سناریوها قابل‌قبول بوده و ضریب همبستگی در بخش آموزش بین ۰/۹۵۹ و ۰/۹۹۹ و در بخش آزمون مابین ۰/۷۵۳ و ۰/۸۷۱

سنجش دقت مدل‌ها

برای بررسی دقت تخمین هر یک از مدل‌ها با استفاده از جذر میانگین مجموع مربع خطاها (RMSE) و متوسط خطای سوء‌گیری (MBE) و ضریب همبستگی (R) و اسنجی انجام گرفت. معادلات مربوط به خطاهای RMSE و MBE به‌صورت زیر می‌باشد:

$$RMSE = \sqrt{\frac{1}{t} \sum_{i=1}^t (P_i - O_i)^2} \quad (۲)$$

$$MBE = \frac{\sum_{i=1}^t (P_i - O_i)}{t} \quad (۳)$$

که در روابط فوق P_i و O_i مقادیر مشاهداتی و تخمینی در زمان i و t تعداد داده‌ها هستند. مقدار RMSE بیانگر متوسط مربع خطا بین مقادیر مشاهداتی و پیش‌بینی شده است و هر چه مقدار آن کمتر باشد نشان‌دهنده دقت بیشتر تخمین مدل است. مقدار MBE نشان‌دهنده مثبت یا منفی بودن خطای محاسباتی بوده و در صورت مثبت یا منفی بودن نمایانگر این مطلب است که مقادیر پیش‌بینی شده به ترتیب بیشتر یا کمتر از مقادیر مشاهداتی هستند. مقدار ضریب همبستگی نیز همیشه بین منفی یک و یک تغییر می‌کند و هر چه از هر دو طرف به صفر نزدیک‌تر شود، نشان‌دهنده تطابق کمتر مقادیر تخمینی و مقادیر اندازه‌گیری شده می‌باشد.

نتایج و بحث

برای پیش‌بینی سطح ایستابی در دشت سراوان از پارامترهای هواشناسی استفاده گردید. در جدول یک مشخصات آماری داده‌های بکار رفته در بخش آموزش و آزمون به تفکیک ارائه شده است. برای

برای سناریوی برتر (سناریوی چهارم) نشان داده شده است. همچنین در شکل ۶ پراکنش داده‌های مشاهداتی در برابر داده‌های برآوردی سطح ایستابی دشت سراوان در بخش آموزش نشان داده شده است که بیشترین مقدار ضریب همبستگی را بین سناریوهای مورد ارزیابی در بخش آموزش داشت.

است. در این مدل همه سناریوها عملکرد خوبی داشته و نتایج نزدیک به هم بوده است اما بر اساس ارزیابی شاخصه‌ای آماری ارائه شده و اولویت‌دهی بخش آزمون سناریوی چهارم به‌عنوان برترین سناریوی مدل MSP معرفی شد. در شکل ۵ پراکنش داده‌های مشاهداتی در مقابل داده‌های برآوردی سطح ایستابی دشت سراوان در بخش آزمون

جدول ۱- مشخصات آماری داده‌های بکار رفته در مدل‌ها

شاخصه‌ای آماری					
مجموعه داده‌ها	متغیرها	میانگین	کمینه	بیشینه	انحراف معیار
آموزش	میانگین دمای حداکثر	۲۹/۶	-۳/۶	۴۱/۷	۸/۳
	میانگین دمای حداقل	۱۵	۱/۱	۲۶/۸	۷/۷
	دمای میانگین	۲۲/۴	۷	۳۴/۳	۷/۷
	میانگین رطوبت نسبی	۲۹/۷	۱۳	۵۸	۱۰
	میزان بارندگی	۸/۷	۰	۹۱/۹	۱۶/۵
	تبخیر ماهانه	۳۰۳/۱	۰	۵۶۵/۵	۱۵۵/۳
	میانگین سرعت باد	۲/۹	۱/۲	۴/۵	-/۶
	مجموع ساعات آفتابی	۲۸۸/۱	۲۱۰	۴۳۲	۳۳/۹
	سطح ایستابی	۲۴	۱۸/۳	۲۹/۴	۲/۶
	آزمون	میانگین دمای حداکثر	۳۰/۳	۱۶/۳	۴۰/۶
میانگین دمای حداقل		۱۵/۲	۰/۴	۲۶/۵	۷/۹
دمای میانگین		۲۲/۸	۸/۳	۳۳/۳	۷/۸
میانگین رطوبت نسبی		۲۷/۸	۱۴	۵۳	۹/۳
میزان بارندگی		۶/۸	۰	۷۹/۴	۱۳/۷
تبخیر ماهانه		۲۸۲/۷	۰	۵۱۱/۲	۱۴۴/۶
میانگین سرعت باد		۳/۵	۲/۲	۳/۴	-/۴
مجموع ساعات آفتابی		۲۹۵/۳	۲۰۳/۳	۳۶۷/۶	۳۷/۵
سطح ایستابی		۲۸/۱	۲۷	۲۹/۳	-/۵

جدول ۲- سناریوهای در نظر گرفته شده جهت مدل‌سازی

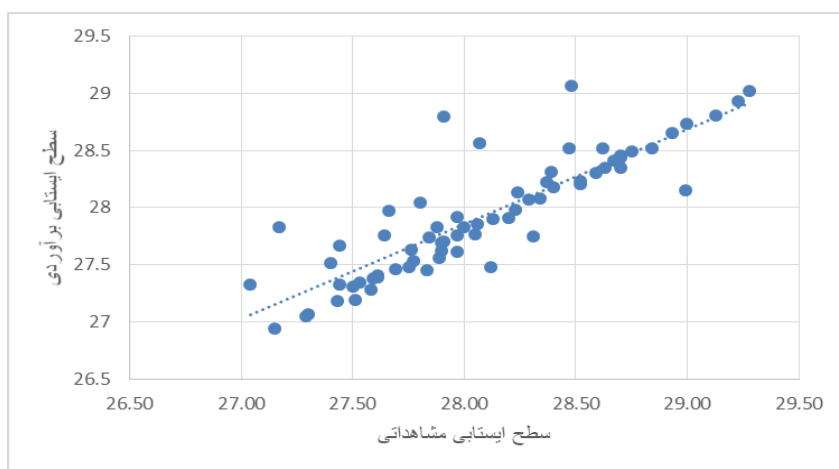
سناریو	متغیرهای ورودی	پارامتر شبیه‌سازی
سناریوی اول	میانگین دمای حداکثر، میانگین دمای حداقل، دمای میانگین، سطح ایستابی ماه قبل	سطح ایستابی
سناریوی دوم	میانگین رطوبت نسبی، میزان بارندگی، تبخیر ماهانه، سطح ایستابی ماه قبل	سطح ایستابی
سناریوی سوم	میانگین سرعت باد، مجموع ساعات آفتابی، سطح ایستابی ماه قبل	سطح ایستابی
سناریوی چهارم	میانگین رطوبت نسبی، میزان بارندگی، تبخیر ماهانه، میانگین سرعت باد، مجموع ساعات آفتابی، سطح ایستابی ماه قبل	سطح ایستابی
سناریوی پنجم	میانگین دمای حداکثر، میانگین دمای حداقل، دمای میانگین، میانگین رطوبت نسبی، میزان بارندگی، تبخیر ماهانه، میانگین سرعت باد، مجموع ساعات آفتابی، سطح ایستابی ماه قبل	سطح ایستابی

برای سناریوی پنجم مدل جنگل تصادفی و سناریوی چهارم مدل MSP در شکل ۷ نشان داده شده است.

به‌منظور مقایسه عملکرد مدل‌سازی دو مدل بکار گرفته شده در بخش آموزش روند تغییرات خطی داده‌های مشاهداتی و برآوردی

جدول ۳- ارزیابی شاخصه‌های آماری سناریوهای مدل RF

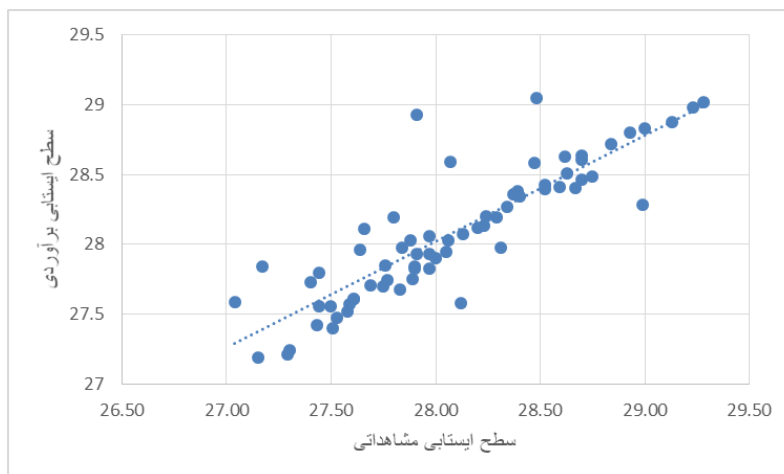
شاخص‌های آماری			سناریو	مجموعه داده‌ها
RMSE	MBE	R		
۰/۸۴۸	-۰/۵۲۷	۰/۹۶۸	سناریوی اول	آموزش
۰/۷۲۹	-۰/۲۳۸	۰/۹۶۶	سناریوی دوم	
۱/۰۱۲	-۰/۷۵۹	۰/۹۶۷	سناریوی سوم	
۰/۸۲۸	-۰/۵۵۴	۰/۹۷۳	سناریوی چهارم	
۰/۶۰۷	-۰/۲۴۳	۰/۹۷۷	سناریوی پنجم	
۱/۶۳۵	-۱/۵۷۳	۰/۸۳۶	سناریوی اول	آزمون
۱/۴۶۳	-۱/۴۰۳	۰/۷۵۳	سناریوی دوم	
۱/۷۸۶	-۱/۷۳۷	۰/۸۵۶	سناریوی سوم	
۰/۳۱۶	-۰/۱۵۷	۰/۸۶۶	سناریوی چهارم	
۰/۳۰۳	-۰/۱۵۱	۰/۸۶۸	سناریوی پنجم	



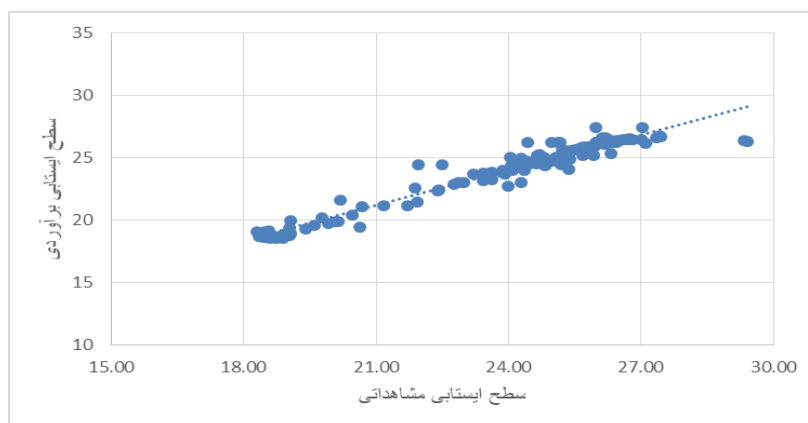
شکل ۴- پراکنش داده‌های مشاهداتی و برآوردی برای سطح ایستابی دشت سراوان در سناریوی پنجم RF

جدول ۴- ارزیابی شاخصه‌های آماری سناریوهای مدل MSP

شاخص‌های آماری			سناریو	مجموعه داده‌ها
RMSE	MBE	R		
۰/۷۵۰	-۰/۰۰۰۱	۰/۹۵۹	سناریوی اول	آموزش
۰/۶۱۳	۰/۰۰۸	۰/۹۷۳	سناریوی دوم	
۰/۲۹۶	۰/۰۱۱	۰/۹۹۰	سناریوی سوم	
۰/۷۲۷	-۰/۰۰۰۱	۰/۹۶۱	سناریوی چهارم	
۰/۷۲۱	-۰/۰۰۱	۰/۹۵۹	سناریوی پنجم	
۰/۳۲۶	-۰/۱۳۷	۰/۸۴۸	سناریوی اول	آزمون
۱/۰۳۶	-۰/۹۶۴	۰/۷۵۳	سناریوی دوم	
۱/۱۸۸	-۱/۱۲۲	۰/۸۶۹	سناریوی سوم	
۰/۲۶۵	۰	۰/۸۷۱	سناریوی چهارم	
۰/۳۸۷	-۰/۲۴۳	۰/۸۳۳	سناریوی پنجم	



شکل ۵- پراکنش داده‌های مشاهداتی و برآوردی سطح ایستابی دشت سراوان در سناریوی برتر M5P

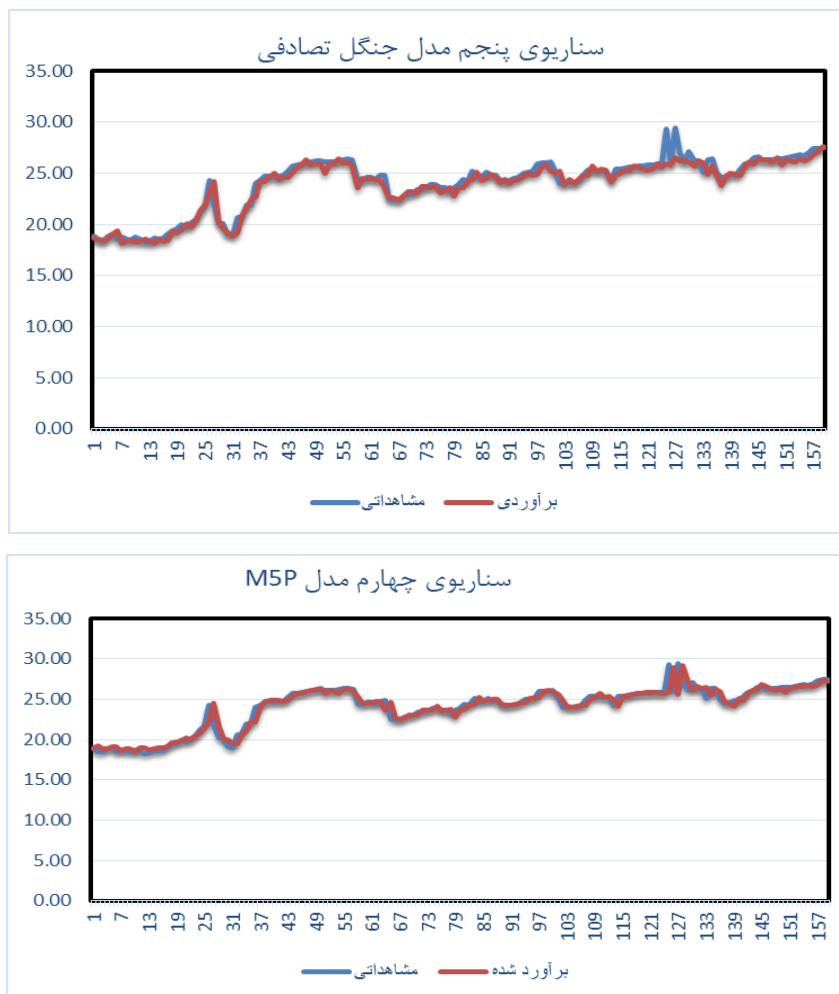


شکل ۶- پراکنش داده‌های مشاهداتی و برآوردی سطح ایستابی دشت سراوان در سناریوی سوم بخش آموزش مدل M5P

نتیجه‌گیری

تاکنون در راستای تخمین و برآورد سطح ایستابی آب‌های زیرزمینی مدل‌های مختلفی طرح شده است که می‌تواند کمک بزرگی در خصوص تصمیمات مدیریت آبخوان‌ها ارائه دهد. با این حال این مدل‌ها در دشت سراوان که به لحاظ آب زیرزمینی در شرایط بحرانی قرار دارد مورد توجه کافی واقع نشده بودند. با توجه به کمبود آب و استفاده بیش‌ازحد از آب‌های زیرزمین، در این پژوهش سطح ایستابی آب زیرزمینی دشت سراوان واقع در استان سیستان و بلوچستان، با استفاده از مدل‌های درختی جنگل تصادفی و M5P مورد ارزیابی قرار گرفت. از داده‌های هواشناسی میانگین دمای حداکثر، میانگین دمای حداقل، دمای میانگین، میانگین رطوبت نسبی، میزان بارندگی، تبخیر ماهانه، میانگین سرعت باد و مجموع ساعت آفتابی و داده آب

زیرزمینی سطح ایستابی ماه قبل به‌عنوان ورودی‌های سناریوهای مختلف برای پیش‌بینی استفاده شد. هر دو مدل مورد استفاده عملکرد مناسبی در برآورد سطح ایستابی دشت سراوان داشتند. در بین سناریوهای مدل‌سازی شده، سناریوی چهارم مدل M5P با مقادیر ضریب همبستگی، MBE و RMSE به ترتیب ۰/۹۶۱، ۰/۰۰۱- و ۰/۷۲۷ در بخش آموزش و ۰/۸۷۱، صفر و ۰/۲۶۵ در بخش آزمون به‌عنوان بهترین مدل در بین سناریوهای دو مدل انتخاب شد. در مجموع نتایج نشان داد که مدل‌های بکار رفته در این تحقیق در زمینه پیش‌بینی سطح ایستابی می‌توانند مؤثر باشند و ابزاری مناسب در خصوص پیاده‌سازی رویکردهای مدیریت آب‌های زیرزمینی می‌باشند.



شکل ۷- مقایسه عملکرد سناریوهای برتر دو مدل در بخش آموزش

زیرزمینی با استفاده از مدل شبکه‌های بیزین پویا مبتنی بر تحلیل حساسیت (مطالعه موردی: دشت بیرجند). مجله پژوهش آب ایران. ۲۹: ۹۱-۱۰۰.

پورصالحی، ف.، خاشعی سیوکی، ع. و هاشمی، س. ر. ۱۴۰۰. 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)، اکوهیدرولوژی، ۸(۴): ۹۷۴-۹۶۱.

حسینی صومه، ۱۳۹۹. مدل‌سازی تغییرات سطح آب زیرزمینی بر اساس روش‌های مبتنی بر هوش مصنوعی (مطالعه موردی: دشت زاوه تربت‌حیدریه)، پژوهشنامه مدیریت حوزه آبخیز، سال یازدهم، شماره ۲۲۳ الی ۲۳۵.

علیمرزائی، ف.، آذرخشی، م.، ملکیان، ا. و رستمی خلج، م. ۱۳۹۸. شبیه‌سازی سطح ایستابی آب زیرزمینی دشت سرخس با ترکیب

نتایج حاصل از این پژوهش نشان داد که مدل درخت تصمیم در دشت سراوان نسبت به مدل جنگل تصادفی عملکرد بهتری داشته است که این مورد با نتایج به‌دست‌آمده از مطالعه پور صالحی و همکاران (۱۴۰۰) تطابق ندارد و دلیل این امر ممکن است مواردی نظیر اختلاف در مشخصات حوضه‌های مورد مطالعه یا پارامترهای هواشناسی مورد استفاده مختلف باشد. همچنین پیشنهاد می‌شود سطح ایستابی دشت سراوان با سایر روش‌های هوش مصنوعی و یادگیری ماشین نیز ارزیابی و با مطالعه حاضر مقایسه شود. استفاده از ترکیب چند روش یادگیری ماشین و داده‌کاوی می‌تواند در جهت افزایش دقت مفید واقع شود و در مدت‌زمان انجام محاسبات مدل صرفه‌جویی کند.

منابع

ابراهیمی ا. روزبهانی ع. و بنجیب م. ا. ۱۳۹۷. پیش‌بینی سطح آب

- HYDROLOGICAL PROCESSES. 23, P. 1437-1443.
- Pourghasemi, H.R. and Kerle, N. 2016. Random forests and evidential belief function-based landslide susceptibility assessment in Western Mazandaran Province, Iran. *Environmental Earth Sciences*, 75(3): 185.
- Quinlan, J.R. 1992. Learning with continuous classes, *Proceedings of the 5th Australian joint Conference on Artificial Intelligence*. Hobart: Singapore.
- Radmanesh F, Golabi M R, Khodabakhshi F, Farzi S, and Zeinali M. 2020. Modeling aquifer hydrograph: Performance review of conceptual MODFLOW and simulator models. *Arabian Journal of Geosciences* 13(5):1-9.
- Rahimikhoob, A. 2014. Comparison between M5 Model Tree and Neural Networks for Estimating Reference Evapotranspiration in an Arid Environment", *Water Resources Management*, 28: 1-13.
- Rahmati O., Pourghasemi H.R., and Melesse A.M. 2016. Application of GIS-based data driven random forest and maximum entropy models for groundwater potential mapping: a case study at Mehran Region, Iran. *Catena*, 137: 360-372.
- Sadeghravesh, M. H., Khosravi, H. and Ghasemian, S. 2015. Application of fuzzy analytical hierarchy process for assessment of combating -desertification alternatives in central Iran. *Natural Hazards*, 75(1), 653 -667.
- Sattari, M.T., Pal, M. Apaydin, H. and Ozturk. F. 2013. M5 model tree application in daily river flow forecasting in Sohu Stream, Turkey. *Water Resour.* 40: 233-242.
- Schapire, R. 1990. The strength of weak learnability. *Journal of Machine learning*, 5, 197-227.
- Senagi, K., Jouandeau, N., and Kamoni, P. 2017. Using parallel Random Forest classifier in predicting land suitability for crop production, *Journal of Agricultural Informatics*, 8(3):23-32.
- Stanley Raj A, Hudson Oliver D, Srinivas Y, Viswanath J. 2017. Wavelet based analysis on rainfall and water table depth forecasting using Neural Networks in Kanyakumari district, Tamil Nadu, India. *Groundwater for Sustainable Development* 5:178-186.
- Sun, Y., Wendi, D., Kim, D.E. and Liong, S.Y. 2015. Application of artificial neural networks in groundwater table forecasting: a case study in Singapore swamp forest. *Hydrology and Earth System Science*. 12: 9317 -9336.
- Wang, Y. and Witten, I.H. 1997. Inducing model trees for continuous classes, *Proceedings of the 9th European Conference on Machine Learning*. Prague, Czech Republic, Springer.
- روش‌های هوش مصنوعی و زمین‌آمار، نشریه پژوهش‌های حفاظت آب و خاک، جلد ۲۶، شماره ۴، صص ۲۰۷ الی ۲۲۲.
- ندیری ع. داداش بابا م. و اصغری مقدم ا. ۱۳۹۷. مدل‌سازی تراز آب زیرزمینی آبخوان دشت تبریز با استفاده از مدل ترکیبی SOM-ANN. *مجله پژوهش آب ایران*. ۳۱: ۹۵-۱۰۲.
- Barzegar R, Fijani E, Asghari Moghaddama A, Tziritis E .2017. Forecasting of groundwater level fluctuations using ensemble hybrid multi-wavelet neural network-based models. *Science of the Total Environment* 599(60):20-31.
- Bonakdar, L. and Etemad-Shahidi, A. 2011. Predicting wave run-up on rubble-mound structures using M5 model tree, *Ocean Engineering*, 38: 111-118.
- Daliakopoulos, I.N., Coulibaly, P. and Tsanis, I.K. 2005. Groundwater level forecasting using artificial neural networks. *Journal of Hydrology*, 309(4): 229 -240.
- Eshghi P. Farzadmehr J. Dasturani MT. Arabs Asadi Z. 2016. The Effectiveness of Intelligent Models in Estimating the River Suspended Sediments (Case Study: Babaaman Basin, Northern Khorasan) *Journal of Watershed Management Research*, 7(14): 88-95.
- Etemad-Shahidi, A. and Mahjoobi, J. 2009. Comparison between M5' model tree and neural networks for prediction of significant wave height in Lake Superior, *Ocean Engineering*, 36: 1175-1181.
- Hand, D.J., Mannila, H. and Smyth, P. 2001. *Principles of data mining*, Cambridge, Mass: The MIT Press.
- He, X.G., Chaney, N.W., Schleiss, M. and Sheffield, J. 2016. Spatial downscaling of precipitation using adaptable random forests, *Water Resour. Res.* 52: 8217-8237.
- Hong, H., Pourghasemi, H.R. and Pourtaghi, Z.S. 2016. Landslide susceptibility assessment in Lianhua County (China): a comparison between a random forest data mining technique and bivariate and multivariate statistical models *Geomorphology*, 259: 105-118.
- Nadiri, A., Chitsazan, N., Tsai, F.T.C. and Asghari Moghaddam, A. 2014. Bayesian artificial intelligence model averaging for hydraulic conductivity estimation. *Journal of Hydrologic Engineering*. 19(3): 520 -532.
- Nasari A. 2018. Comparison of the application of fourteen temporal series patterns to analyze and predict changes in groundwater level in Marand plain (northern margin of Urmia Lake). *Irrigation and Drainage* 13(1):58-68.
- Pal, M., Deswal, S. 2009. M5 model tree based modelling of reference evapotranspiration:

networks, *Environmental Modelling & Software*, 55: 107-119.

Wolfs, V. and Willems, P. 2014. Development of discharge-stage curves affected by hysteresis using time varying models, model trees and neural

Evaluating the ability of random forest and M5P tree-based models in groundwater level estimating of Saravan plain using meteorological data

S Sepahian*¹, SA Hashemi Monfared ², AA Qaderi³

Received: Oct.15, 2023

Accepted: Dec.29, 2023

Abstract

Groundwater is one of the key sources of water on the surface of the earth, which is important in agriculture, industry and drinking, and is considered as the guarantor of life for the future generation. In this research, 227 data series related to meteorological and groundwater parameters of Saravan Plain in Sistan and Baluchistan province were used to predict and estimate the groundwater level using intelligent machine learning models. Out of this number of data series, 159 series were allocated for use in model training and 68 series for model testing in designed scenarios. The performance of random forest and M5P models was evaluated in estimating the monthly mean groundwater level of Saravan Plain. The performance of the models was compared using statistical indices of correlation coefficient (R), root mean square error (RMSE) and mean bias error (MBE). The obtained results showed that both used models have a good ability for groundwater level modeling, but the M5P model with a higher correlation coefficient has fewer errors than the random forest model. The fourth scenario of the M5P model is selected as the best model among the scenarios of two models with correlation coefficient, MBE and RMSE values of 0.961, -0.0001 and 0.727 in the training section and 0.871, zero and 0.265 in the test section, respectively. It is suggested that the groundwater level of Saravan Plain be evaluated with other methods of artificial intelligence and machine learning, and compared with the present study.

Keywords: Aquifer, Groundwater Level, Decision Tree, Random Forest, Water Resources Management

1- PhD student, Department of civil engineering, Shahid Nikbakht Faculty of Engineering, University of Sistan and Baluchestan

2- Associate Professor, Department of civil engineering, Shahid Nikbakht Faculty of Engineering, University of Sistan and Baluchestan

3- Assistant Professor, Department of civil engineering, Shahid Nikbakht Faculty of Engineering, University of Sistan and Baluchestan

(*- Corresponding Author Email: s.sepaheyan@gmail.com)