

مقایسه و کاربرد مدل‌های جنگل تصادفی، چاید و زمین‌آمار در پیش‌بینی هدایت هیدرولیکی اشباع خاک

حلیمه پیری^{۱*}، مجتبی مبارکی^۲، معصومه میر^۳

۱. دانشیار، گروه مهندسی آب، دانشکده آب و خاک، دانشگاه زابل

۲. دانش‌آموخته کارشناسی ارشد، گروه مهندسی آب، دانشکده آب و خاک، دانشگاه زابل

۳. دانشجوی کارشناسی، گروه مهندسی آب، دانشکده آب و خاک، دانشگاه زابل

(تاریخ دریافت ۱۴۰۱/۱۰/۱۱؛ تاریخ بازنگری ۱۴۰۱/۱۱/۱۱ تاریخ تصویب ۱۴۰۱/۱۲/۱۲)

چکیده

هدایت هیدرولیکی اشباع یکی از خصوصیات مهم هیدرولیکی در علوم مرتبط با آب، خاک و کشاورزی است که در مدل‌سازی حرکت املاح و آب در خاک بسیار اهمیت دارد. اندازه‌گیری آزمایشگاهی و صحرایی آن دشوار، وقت‌گیر و پرهزینه است و امکان شناسایی تغییرپذیری مکانی و زمانی آن در مقیاس وسیع وجود ندارد. این پارامتر را می‌توان با استفاده از پارامترهای زودیافت خاک برآورد کرد. تحقیق حاضر به منظور پیش‌بینی هدایت هیدرولیکی اشباع خاک با استفاده از روش‌های جنگل تصادفی، چاید و زمین‌آمار در شهرستان هیرمند انجام گرفت. برای این منظور، تعداد ۱۳۰ نمونه خاک از سطح (۰-۳۰ سانتی‌متری) جمع‌آوری و برای انجام آزمایش و آنالیزها، به آزمایشگاه انتقال داده شد. در آزمایشگاه پارامترهای هدایت هیدرولیکی اشباع خاک، بافت خاک، کربن آلی، اسیدیته، هدایت الکتریکی و کربنات کلسیم اندازه‌گیری شد. سپس با استفاده از پارامترهای اندازه‌گیری شده و با استفاده از مدل‌های چاید، جنگل تصادفی و زمین‌آمار هدایت هیدرولیکی اشباع برآورد شد. ۲۰ الگوی مختلف از ترکیب پارامترهای زودیافت خاک به عنوان ورودی به مدل چاید و جنگل تصادفی در نظر گرفته شد. از بین ترکیبات مختلف بر اساس MAE کمتر و R بالاتر ترکیب بهتر انتخاب شد. نتایج نشان داد مدل جنگل تصادفی با داشتن بالاترین ضریب تبیین (۰/۹۸) و کمترین مقدار خطای مطلق میانگین (و ۰/۰۰۱۹) بهترین مدل برای پیش‌بینی هدایت هیدرولیکی اشباع خاک در منطقه هیرمند است. نتایج پهنه‌بندی نشان داد مقدار هدایت هیدرولیکی اشباع خاک در نقاط غرب، مرکز و شمال شرق بیشتر از سایر نقاط بود.

کلمات کلیدی: توابع انتقالی، جنگل تصادفی، روند هدایت هیدرولیکی، ویژگی‌های زودیافت خاک.

مقدمه

هدایت هیدرولیکی اشباع یکی از خصوصیات مهم هیدرولوژیکی خاک برای تخمین مدل‌های هیدرولوژیکی و تعیین فاصله مناسب زهکش‌ها و حرکت املاح در خاک است که تغییرات مکانی و الگوی پراکنش آن اغلب ضروری است. تعیین هدایت هیدرولیکی اشباع به طور مستقیم از طریق روش‌های آزمایشگاهی و صحرایی انجام می‌شود. اندازه‌گیری با هر دو روش مزرعه‌ای و آزمایشگاهی دشوار، وقت‌گیر و پرهزینه است. علاوه بر این، ممکن است نتایج آن با توجه به تغییرات زمانی و مکانی در خواص فیزیکی و هیدرولیکی خاک دقیق نباشد. این امر به توسعه و استفاده گسترده از روش‌های غیرمستقیم منجر شده است [۱]. روش غیرمستقیم به دست آوردن داده‌های دیرپافت خاک، استفاده از توابع انتقالی است [۲]. به دلیل اینکه توابع انتقالی خصوصیات دیرپافت خاک را با استفاده از خصوصیات زودپافت آن برآورد می‌کنند، باعث صرفه‌جویی در وقت و هزینه می‌شوند [۳].

از جمله مدل‌های برآوردکننده توابع انتقالی شامل مدل‌های رگرسیونی و شبکه‌های عصبی مصنوعی هستند. مدل‌های آماری یا معادلات رگرسیونی تابعی از فرضیات و ملاحظات ذاتی آزمایش‌ها در تحلیل است. در روش‌های آماری واقعیت پدیده به کمک یک معادله جبری بیان می‌شود که فرضیات اصلی مسئله در آن لحاظ شده است. از آنجا که برای پیش‌بینی پارامترهای هیدرولیکی در خاک، چندین متغیر ورودی و خروجی مؤثرند، تحلیل آماری این نوع مسائل مستلزم ارائه تعدادی روابط ریاضی است که اغلب برای استفاده و تفسیر دشوار است. اما امروزه با توسعه سریع فناوری‌های پردازش رایانه‌ای و ایجاد نرم‌افزارهای مربوطه، از مزایای فناوری هوش مصنوعی مانند شبکه‌های عصبی برای مسائل مربوط به مدل‌سازی سیستم‌ها و فرایندها استفاده می‌شود [۴]. محققان با استفاده از پارامترهایی همچون تخلخل مؤثر، فراوانی نسبی ذرات خاک، ماده آلی و چگالی ظاهری، توابعی برای برآورد هدایت آبی اشباع خاک ارائه کردند [۵ و ۶]. روش‌های شبکه عصبی مصنوعی، ماشین بردار پشتیبان و درخت تصمیم برای پیش‌بینی هدایت هیدرولیکی اشباع خاک مورد مقایسه قرار گرفت. نتایج حاصل بیانگر عملکرد بسیار مطلوب روش‌های هوش مصنوعی در پیش‌بینی هدایت هیدرولیکی اشباع خاک است [۷].

در شرایطی که امکان اندازه‌گیری هدایت هیدرولیکی امکان‌پذیر نباشد، می‌توان با استفاده از پارامترهای زودپافت خاک و روش‌های هوش مصنوعی، هدایت هیدرولیکی را تخمین زد. مدل‌هایی که تا کنون برای ارزیابی هدایت هیدرولیکی به کار رفته‌اند، به طور نسبی از داده‌های کمتری استفاده کرده‌اند. روش‌های جدید با بهره‌گیری از رگرسیون‌های گروهی ارائه شده و در حال توسعه هستند [۸]. جنگل تصادفی^۱ (RF) یکی از انواع روش‌های یادگیری است که از الگوریتم‌های پایه برای پیش‌بینی‌های چندگانه تکراری استفاده می‌کند [۱۰]. جنگل تصادفی اهمیت نسبی متغیرها را ارزیابی می‌کند و قادر به انتخاب متغیرهای مهم نیز است و در عین حال پارامترسازی آن نسبت به روش‌های دیگر مانند شبکه‌های عصبی، محاسبات ساده‌تری دارد [۱۱]. در تحقیقی قابلیت انتقال آبخوان دشت ملکان با استفاده از شبکه عصبی مصنوعی، منطق فازی و جنگل تصادفی مورد بررسی قرار گرفت. بر اساس نتایج از بین مدل‌های بررسی‌شده، مدل جنگل تصادفی دقت و توانایی بیشتری در شبیه‌سازی داشته است [۱۲]. از ماشین بردار پشتیبان، چاید و جنگل تصادفی برای برآورد تبخیر و تعرق روزانه دشت سیستان استفاده شد. نتیجه تحقیق یادشده نشان داد جنگل تصادفی برآورد بهتری از تبخیر و تعرق روزانه داشت [۱۳]. از دیگر روش‌های مرسوم برای برآورد خصوصیات فیزیکی و شیمیایی خاک بر اساس تعداد محدودی داده اندازه‌گیری‌شده، روش‌های میان‌یابی زمین‌آماري است. زمین‌آمار پیوستگی طبیعی خاک را مشخص می‌کند و واریانس تصادفی در مدل‌سازی مکانی خاک را در همبستگی با سایر خصوصیات خاک نشان می‌دهد [۱۴]. تحقیقات زیادی در خصوص استفاده از زمین‌آمار در پیش‌بینی هدایت هیدرولیکی خاک و سایر پارامترهای وابسته به خاک ارائه شده است. مدل‌سازی زمانی و مکانی سطح آب زیرزمینی دشتستان بوشهر با استفاده از هوش مصنوعی و زمین‌آمار انجام شد. نتایج پژوهش یادشده نشان داد روش شبکه عصبی مصنوعی با داشتن R^2 بالاتری (۰/۹۹۳)، کمترین مجذور میانگین خطا (۰/۲۹) و کمترین میانگین مطلق خطا (۰/۰۲۴) به عنوان روش برتر برای پیش‌بینی سطح ایستابی انتخاب شد [۱۵].

می‌توان آثار شوره را به‌وضوح در سطح خاک‌های منطقه مشاهده کرد. بادهای ۱۲۰ روزه در این منطقه منحصربه‌فرد بوده و یکی از عوامل دخیل مهم در مقدار بالای تبخیر و تعرق است. فرسایش بادی یکی از چالش‌های این منطقه است که اثرات منفی بر محیط زیست و تأسیسات زیربنایی آن دارد. کاربری زمین‌های منطقه شامل کشاورزی، مرتع و همچنین زمین‌های غیر قابل استفاده هستند. عمق خاک منطقه پایین و خاک آن ماحصل رسوبات ته‌نشین شده ناشی از سیلاب‌های ورودی از کشور افغانستان است [۱۵]. در اراضی تحت کشت معمولاً گندم، جو و یونجه به صورت غرقابی کشت می‌شوند که منابع آب آن‌ها از رودخانه هیرمند، چاه‌نیمه‌ها و آب زیرزمینی در منطقه تأمین می‌شود. شکل ۱ موقعیت منطقه مورد مطالعه و نقاط نمونه‌برداری شده را نشان می‌دهد.

برای انجام کار تعداد ۱۳۰ نمونه خاک از عمق ۰-۳۰ سانتی‌متری خاک بر اساس شبکه‌بندی منظم با ابعاد ۱۵۰۰×۱۵۰۰ m برداشت شد. نمونه‌های برداشت شده به آزمایشگاه منتقل شد. در آزمایشگاه نمونه‌ها هواخشک و سپس از الک ۲ میلی‌متری عبور داده شدند. در آزمایشگاه بافت خاک به روش هیدرومتری [۱۶]، کربن آلی به روش والکی-بلک [۱۷]، درصد آهک به روش تیتراسیون با اسید [۱۸]، هدایت الکتریکی با استفاده از دستگاه هدایت الکتریکی سنج [۹]، pH خاک با pH متر [۱۹] و هدایت هیدرولیکی اشباع به روش بار ثابت اندازه‌گیری شد.

الگوریتم چاید^۱ (CHAID)

چاید مخفف مجذور مربع شناسایی‌کننده تعاملات خودکار است. این روش آماری بسیار مؤثر برای بخش‌بندی و رشد درخت، به وسیله کاس توسعه یافته است. با استفاده از مفهوم تست آماری به عنوان یک معیار، چاید تمام مقادیر صفت پیش‌بینی‌کننده بالقوه را ارزیابی می‌کند. این الگوریتم مقادیری را که به صورت آماری همگن در نظر گرفته شده‌اند با توجه به متغیر هدف و حفظ تمام مقادیری که ناهمگن هستند، ادغام می‌کند [۲۰].

این الگوریتم بهترین پیش‌بینی‌کننده را از شاخه درخت تصمیم انتخاب می‌کند، به طوری که هر فرزندگره از یک گروه از ارزش‌های همگون نسبت به صفت انتخاب‌شده

تحقیقی دیگر روش‌های هوش مصنوعی و زمین‌آمار در پیش‌بینی شوری سطحی خاک مورد مقایسه قرار گرفت. نتایج پژوهش یادشده نشان داد روش‌های هوش مصنوعی نسبت به روش زمین‌آمار شوری خاک را بهتر برآورد می‌کنند [۱۶].

همان‌طور که ملاحظه شد، در تحقیقات انجام‌شده از روش‌های مختلفی برای پیش‌بینی هدایت هیدرولیکی اشباع خاک استفاده شده است. با توجه به اینکه همه روش‌ها در هر منطقه کارایی ندارد و از طرف دیگر مقایسه بین روش‌های هوش مصنوعی و زمین‌آمار در بررسی هدایت هیدرولیکی خاک نشده است، در این تحقیق به بررسی برآورد هدایت هیدرولیکی اشباع خاک دشت قرقری از توابع شهرستان هیرمند با استفاده از زمین‌آمار و روش‌های جنگل تصادفی و چاید پرداخته شد. با توجه به اهمیت هدایت هیدرولیکی در طرح‌های آبیاری و زهکشی و اجرای طرح آبیاری ۴۶ هزار هکتاری دشت سیستان و بالا بودن هزینه اندازه‌گیری صحرائی این پارامتر، لازم است با استفاده از داده‌های زودیافت خاک و روش‌های هوش مصنوعی، به شبیه‌سازی این پارامتر در منطقه مورد مطالعه پرداخته شود. همچنین با توجه به اینکه تغییرات مکانی در ویژگی‌های خاک امری معمول است، ولی شناخت این تغییرات به‌ویژه در اراضی کشاورزی برای برنامه‌ریزی دقیق و مدیریت امری اجتناب‌ناپذیر است و آگاهی از این مسئله برای بهبود در سودآوری و نیل به بهره‌برداری پایدار ضرورت دارد [۱۷]، پهنه‌بندی هدایت هیدرولیکی خاک منطقه با استفاده از GIS انجام شد.

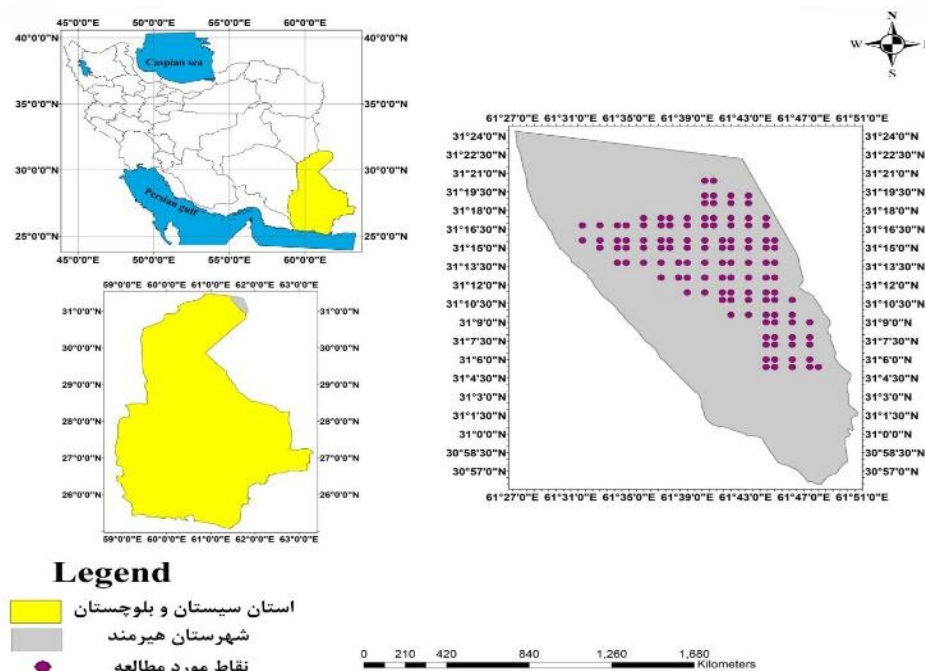
مواد و روش‌ها

منطقه مطالعه‌شده

بخش قرقری از توابع شهرستان هیرمند واقع‌شده در شمال استان سیستان و بلوچستان است که از لحاظ قرارگیری جغرافیایی نسبت به شهرستان زابل در ۳۳ کیلومتری شمال شرقی شهرستان زابل به طول جغرافیایی 0° 45' 61 تا 0° 30' 61 شرقی و عرض جغرافیایی 0° 31' 31 تا 0° شمالی واقع شده است. متوسط بارندگی سالیانه آن ۵۰ mm و ارتفاع از سطح دریا به طور متوسط ۴۷۵ m است. تبخیر و تعرق سالانه در این منطقه ۴۵۰۰ mm تا ۵۰۰۰ mm است که بسیار زیاد است، به همین دلیل

می‌تواند بیش از دو دسته در هر سطح خاص از درخت ایجاد کند. بنابراین گرایش به ساخت درخت نسبت به روش رشد باینری گسترده‌تر است. این الگوریتم برای تمام انواع متغیرها پاسخ‌گو است و دو متغیر وزنی و فراوانی را می‌پذیرد. مدل یادشده همچنین مقادیر مفقودی را با یک طبقه‌بندی معتبر ایجاد می‌کند [۲۱].

هستند. این فرایند به صورت بازگشتی تا آنجا که درخت به طور کامل رشد می‌کند، ادامه می‌یابد. آزمون آماری مورد استفاده بستگی به سطح اندازه‌گیری میدان هدف دارد. اگر صفت هدف پیوسته باشد، یک آزمون F صورت می‌گیرد. اگر صفت هدف دسته‌ای باشد، یک آزمون مجذور مربع به کار برده می‌شود. چاید یک روش باینری نیست و



شکل ۱. موقعیت نقاط در منطقه مورد مطالعه

روش یک تکنیک مدل ناپارامتری و متعلق به خانواده روش‌های دسته‌جمعی است [۲۴]. جنگل تصادفی با استفاده از مجموعه‌ای از درخت‌ها با در نظر گرفتن n داده مشاهده‌ای مستقل ساخته می‌شود:

$$(Y_i, X_i), i = 1, \dots, n \quad (1)$$

این الگوریتم از جمله دسته‌بندی‌هایی است که روش Bagging را استفاده می‌کند. این روش ترکیبی از چندین درخت تصمیم است که در ساخت آن چندین نمونه بوت استرپ از داده‌ها شرکت دارند و در ساخت هر درخت به طور تصادفی تعدادی از متغیرهای ورودی شرکت می‌کنند. با استفاده از روش بوت استرپ به تعداد زیاد (به عنوان مثال ۲۰۰۰ مرتبه) نمونه‌های n تایی از مجموعه داده‌های مشاهداتی اولیه، نمونه‌برداری همراه با جای‌گذاری می‌شوند. طی فرایند نمونه‌گیری حدود یک سوم از داده‌ها نمونه‌گیری نمی‌شوند و به

جنگل تصادفی^۱ (RF)

جنگل تصادفی (RF) یک روش گروهی است که چند الگوریتم درختی را برای تولید یک پیش‌بینی مکرر از هر پدیده ترکیب می‌کند [۲۲]. 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 نبود توزیع نرمال داده‌های استفاده‌شده، در تجزیه و تحلیل بگنجانند و ترکیب کند. مدل جنگل تصادفی برای بسیاری از مجموعه داده‌ها، دسته‌بندی را با سرعت زیادی انجام می‌دهد و به خلاف مدل‌های کلاسیک همچون رگرسیون که تنها بر پایه یک مدل تکیه دارند، با استفاده از صدها و هزاران درخت از اطلاعات بیشتری در داده‌ها استفاده می‌کند تا بتوان استنباط بهتری از متغیرها داشت. این

عنوان نمونه خارج از کیسه (از این داده‌ها برای تعیین متغیرهای مهم و همچنین، برآورد نارایب خطا استفاده می‌شود) در نظر گرفته می‌شوند. سپس روی هر نمونه بوت استرپ یک درخت گسترش داده می‌شود. طی فرایند ساخت درخت در هر شاخه، از بین تمام M متغیر مستقل به صورت تصادفی m متغیر برای تقسیم شدن انتخاب می‌شود. برای حالت رگرسیونی نسبت m/M برابر یک سوم است و برای کلاسه‌بندی برابر با $m = \sqrt{M}$ پیشنهاد شده است. پس از ساخت تمام درخت داده‌های تست به درخت معرفی می‌شود و به تعداد درخت‌ها برای بردار ورودی یک خروجی به دست می‌آید. با میانگین‌گیری این خروجی‌ها، خروجی نهایی مدل و در نظر گرفتن توزیع تجربی خروجی‌ها مقادیر صدک‌ها و دامنه عدم قطعیت محاسبه می‌شود. روش جنگل تصادفی زمانی که تعداد مشاهدات در مقایسه با تعداد پیش‌بینی‌کننده‌ها نسبتاً کم باشد، یک روش پیش‌بینی کارآمد است [۱۰].

زمین‌آمار

اساسی‌ترین ابزار در زمین‌آمار نیم‌تغییرنا است که برای تشریح و توصیف ارتباط مکانی یک متغیر به کار می‌رود. نیم‌تغییرنا کمیتی برداری است که درجه همبستگی مکانی و شباهت بین نقاط اندازه‌گیری شده را بر حسب مربع تفاضل مقدار دو نقطه و با توجه به جهت و فاصله آن‌ها برآورد می‌کند. از مهم‌ترین ویژگی‌های نیم‌تغییرنا می‌توان به اثر قطعه‌ای^۱، آستانه^۲ و دامنه تأثیر^۳ اشاره کرد [۲۳]. به منظور بررسی روابط مکانی بین نقاط اندازه‌گیری شده از مدل‌سازی واریوگرام یا نیم‌تغییرنا استفاده شد. سپس با استفاده از روش کریجینگ به میان‌یابی و پیش‌بینی داده‌ها پرداخته شد. کریجینگ یکی از روش‌های متداول میان‌یابی زمین‌آمار است که بهترین تخمینگر خطی نارایب نام گرفته است. این روش با استفاده از مقادیر معلوم و یک نیم‌تغییرنا، مقادیر متغیر در نقاط فاقد آمار را برآورد می‌کند. در عمل نیم‌تغییرنمای تجربی که با نماد $\gamma^*(h)$ نشان داده می‌شود، که با استفاده از رابطه ۲ محاسبه شد:

$$\gamma^*(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} [Z(x_i) - Z(x_i + h)]^2 \quad (2)$$

1. Nugget Structure
2. Range
3. Sill

$$Z^*(x_0) = \sum_{i=1}^n \lambda_i Z(x_i) \quad \text{with} \quad \sum_{i=1}^n \lambda_i = 1 \quad (3)$$

که در آن $Z^*(x_0)$: مقدار تخمین زده شده متغیر Z در نقطه‌ای به مختصات x_0 ; مقدار مشاهده شده متغیر Z در نقطه‌ای به مختصات x_i ; λ_i : وزن یا اهمیت نسبت داده شده به متغیر Z در نقطه‌ای به مختصات x_i و n : تعداد نقاط همسایگی است.

برای انجام محاسبات و رسم واریوگرام‌های مربوط به زمین‌آمار از نرم‌افزار GS^+ استفاده شد. نرم‌افزار GS^+ یک برنامه تحلیل زمین‌آمار و تهیه نقشه است که کمک می‌کند تا روابط مکانی موجود در داده‌های ژئورفرنس شده را سریع و به طور کارآمدی استخراج کند. نرم‌افزار GS^+ داده‌های مکانی را از لحاظ روابط درونی آن‌ها تحلیل کرده و سپس از این اطلاعات استفاده می‌کند تا بهترین و از لحاظ آماری قوی‌ترین نقشه‌های ممکن را برای منطقه نمونه برداری شده ارائه کند [۲۵].

متغیرهای ورودی به مدل‌ها

با توجه به اینکه هدف از تحقیق برآورد هدایت هیدرولیکی اشباع با استفاده از داده‌های زودیافت خاک است، داده‌های ورودی به مدل‌ها الگوهای مختلفی از ترکیب داده بافت خاک، pH، EC، درصد آهک و کربن آلی خاک است. الگوها در جدول ۱ آورده شده است.

جدول ۱. ترکیبات مختلف ورودی به مدل‌های مورد استفاده در تحقیق

تعداد متغیر	تعداد پارامترهای ورودی به مدل	سناریو
۱	بافت خاک	S1
۲	بافت خاک، pH	S2
۳	بافت خاک، pH، EC	S3
۴	بافت خاک، pH، EC، درصد آهک	S4
۵	بافت خاک، pH، EC، درصد آهک، درصد کربن آلی خاک	S5
۴	pH، EC، درصد آهک، درصد کربن آلی خاک	S6
۳	EC، درصد آهک، درصد کربن آلی خاک	S7
۲	درصد آهک، درصد کربن آلی خاک	S8
۱	درصد کربن آلی خاک	S9
۱	درصد آهک	S10
۴	بافت خاک، pH، EC، درصد کربن آلی خاک	S11
۴	بافت خاک، pH، درصد آهک، درصد کربن آلی خاک	S12
۴	بافت خاک، EC، درصد آهک، درصد کربن آلی خاک	S13
۲	EC، pH	S14
۲	بافت خاک، درصد کربن آلی خاک	S15
۳	بافت خاک، درصد آهک، درصد کربن آلی خاک	S16
۴	بافت خاک، pH، EC، درصد کربن آلی خاک	S17
۴	بافت خاک، pH، درصد آهک، کربن ماده آلی خاک	S18
۴	بافت خاک، EC، درصد آهک، درصد کربن آلی خاک	S19
۲	بافت خاک، درصد آهک	S20

مقایسه بین آن‌ها، روش برتر در برآورد هدایت هیدرولیکی اشباع خاک در منطقه مورد مطالعه انتخاب می‌شود.

معیارهای ارزیابی مدل

برای ارزیابی صحت و اعتبار مدل‌های مورد استفاده از شاخص‌های آماری ضریب تبیین (R^2) و میانگین مطلق خطا (MAE) استفاده شد (روابط ۴ و ۵).

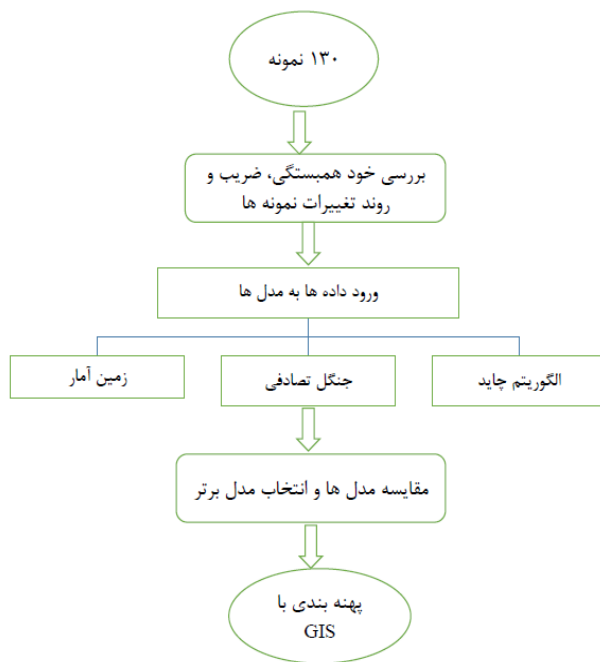
$$R^2 = 1 - \frac{\sum_{i=1}^n (x_i - x_i^*)^2}{\sum_{i=1}^n (x_i - \bar{x}_i)^2} \quad (4)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |x_i - x_i^*| \quad (5)$$

در این روابط x_i مقدار اندازه‌گیری شده، x_i^* مقادیر برآورد شده و n تعداد نمونه است.

شکل ۲ مراحل انجام تحقیق را نشان می‌دهد.

برای انتخاب مناسب‌ترین مدل نفوذ، بر اساس معیارهای خطای یادشده، از روش رتبه‌بندی برای داده‌های تست استفاده می‌شود. به این معنا که به هر یک از مدل‌ها بر اساس معیارهای یادشده رتبه‌ای داده می‌شود و در پایان پس از مقایسه مجموع رتبه‌های به‌دست‌آمده، مناسب‌ترین مدل برای تخمین بهترین معادله انتخاب می‌شود. شیوه رتبه‌بندی این مدل‌ها به این صورت است که در هر روش، مدلی که کمترین میزان خطای MAE را دارد، پایین‌ترین رتبه، یعنی رتبه ۱ و به مدل‌ها یا ترکیبات ورودی دیگر در هر یک از دسته‌های ورودی بر حسب میزان معیار خطاهای یادشده، رتبه‌های ۲ تا آخر تعلق می‌گیرد. اما در مورد معیارهای ضریب همبستگی (R) مدل یا ترکیبی که بیشترین مقدار این معیارها را دارد، رتبه ۱ اختصاص می‌یابد. همچنین به مدلی که میزان معیار خطای آن‌ها یکسان باشند، رتبه یکسان تعلق می‌گیرد. در نهایت پس از انتخاب مدل‌ها یا ترکیبات مناسب هر روش و



شکل ۲. الگوریتم مورد استفاده در تحقیق

جدول ۲. توصیف آماری خصوصیات فیزیکی و شیمیایی نمونه‌های خاک

چولگی	ضریب تغییرات	انحراف معیار	میانگین	میانگین	حداکثر	حداقل	واحدها	خصوصیات خاک
۰/۶۸	۲۷/۰۱	۵/۷۷	۲۰	۲۱/۲۹	۳۷/۵	۸	درصد	آهک
۱/۷۹	۹۴/۲۴	۱۴/۲۸	۹/۸۲	۱۵/۱۵	۶۴	۲/۴۲	ds/m	هدایت الکتریکی
۰/۲۳	۶/۷۲	۰/۵۳	۸	۷/۹۶	۱۰/۲۲	۶/۷۰	(-log[H+])	اسیدیته خاک
۰/۹۹	۷۳/۸۷	۳/۴۳	۳/۵	۴/۶۵	۱۷/۵	۰/۰۰	درصد	رس
۰/۴۱	۷/۰۷	۵/۸۶	۸۲	۸۲/۸۹	۹۶	۷۳	درصد	شن
-۰/۱۹	۳۷/۲۶	۴/۶۳	۱۲/۵	۱۲/۴۴	۲۱	۱	درصد	سیلت
۱/۰۷	۴۷/۸۳	۰/۲۱	۰/۴۱	۰/۴۴	۱/۳۴	۰/۱۱	درصد	کربن آلی
۰/۱۳	۱۷/۸۷	۰/۰۷۲	۰/۰۴۱	۰/۰۴	۰/۰۵۷	۰/۰۲۵	m/day	هدایت هیدرولیکی اشباع

نتایج و بحث

متقابل عوامل مدیریتی، مانند مصرف کود و عوامل ذاتی مانند وضعیت زهکشی در خاک‌ها باشد [۲۷]. ضریب تغییرات بالای هدایت الکتریکی (۹۴/۲۴) در منطقه مورد مطالعه می‌تواند ناشی از بالا بودن سطح آب‌های زیرزمینی و شدت بالای تبخیر ۴۵۰۰ تا ۵۰۰۰ میلی‌متر و همچنین ناشی از کنده شدن و به جا ماندن رسوبات بادرفتی سطح دریاچه هامون باشد که در سال‌های اخیر خشک شده و باعث به جا ماندن نمک‌های محلول در سطح خاک منطقه مورد مطالعه شده است. تغییرپذیری در ویژگی‌های خاک نتیجه تغییر در محیط‌های رسوب‌گذاری و یا اختلاف در مراحل خاک‌سازی یا هیدرولوژیکی در شکل‌های اراضی مختلف است. تغییرات بالای هدایت الکتریکی در خاک ممکن است به دلیل بالا بودن سطح آب زیرزمینی و

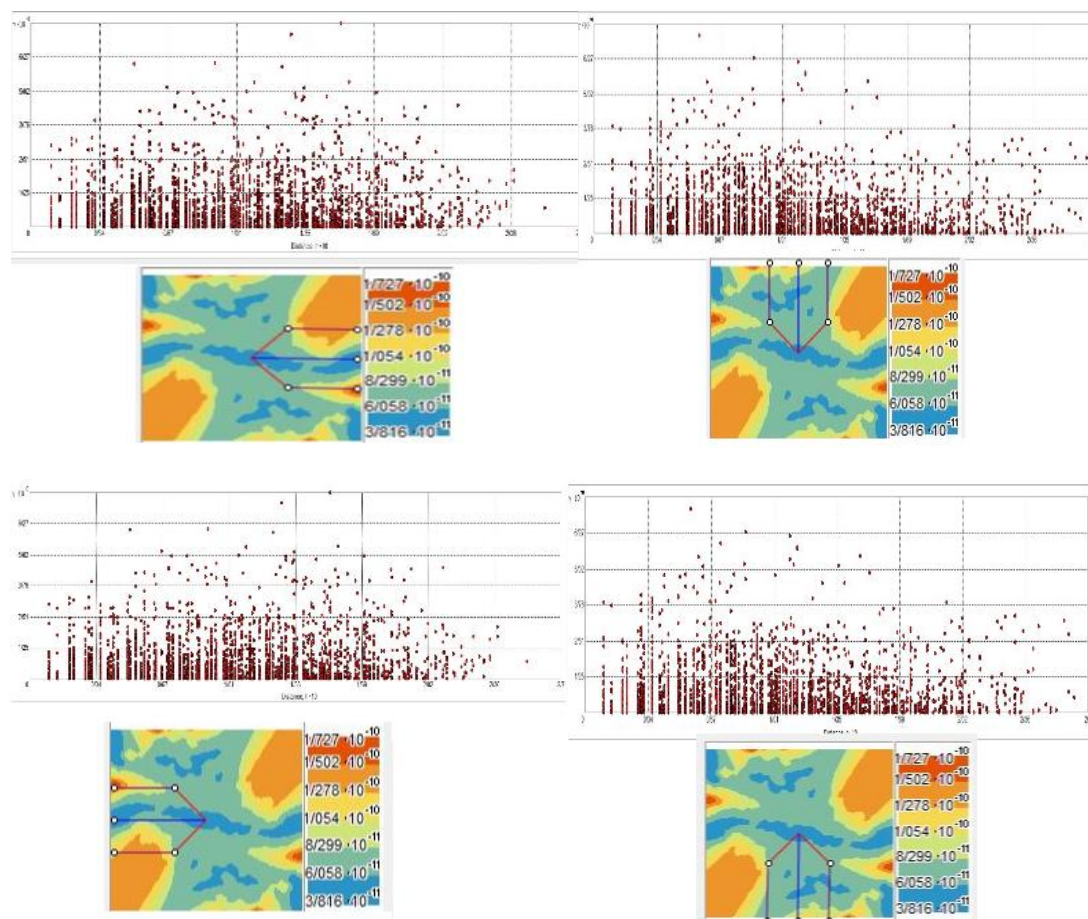
برخی نمایه‌های آمار توصیفی پارامترهای فیزیکی و شیمیایی خاک مورد مطالعه در جدول ۲ آورده شده است. ضریب تغییرات (پراکندگی) شاخصی است که برای اندازه‌گیری توزیع پراکندگی داده‌های آماری و همچنین مقایسه پراکندگی دو یا چند متغیر به کار می‌رود. در این مطالعه کمترین مقدار ضریب تغییرات (۶/۷۲ درصد) برای اسیدیته خاک و بیشترین مقدار آن (۹۴/۲۴ درصد) برای هدایت الکتریکی خاک است. کمترین ضریب تغییرات برای اسیدیته خاک گزارش شده است [۲۶]. محققان بیان داشته‌اند که ضریب تغییرات کم برای اسیدیته متأثر از عوامل ذاتی همچون مواد مادری در رفتار این ویژگی است. در حالی که ضریب تغییرات زیاد می‌تواند ناشی از اثر

هیدرولیکی اشباع خاک را در چهار جهت جغرافیایی نشان می‌دهد. همان‌طور که در شکل ۳ مشاهده می‌شود، همبستگی بین مقادیر پایین‌تر هدایت هیدرولیکی اشباع بیشتر از مقادیر بالاتر بوده است.

حرکت موئینگی آب به سمت بالا و تجمع املاح در بعضی از افق‌های سطحی خاک باشد [۲۸].

خودهمبستگی بین داده‌ها

در شکل ۳ خودهمبستگی سمی‌واریوگرام داده‌های هدایت



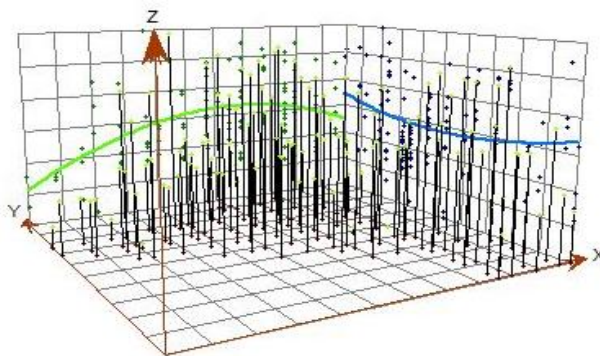
شکل ۳. نمودار خودهمبستگی سمی‌واریوگرام هدایت هیدرولیکی اشباع خاک جهت‌های شمال، جنوب، شرق و غرب

منفی وجود دارد [۱۳]. مطابق تحقیقات انجام‌شده در منطقه مورد مطالعه هدایت الکتریکی از غرب به شرق منطقه ابتدا کاهش و سپس افزایش و از شمال به جنوب روند کاهشی داشته است [۱۳]. این نتایج با یافته‌های سایر محققان همخوانی دارد. ایشان بیان داشتند هدایت الکتریکی بالا باعث تخریب و ناپایداری ساختمان خاک می‌شود [۳۰ و ۲۹].

جدول ۳ نتایج ضریب تبیین و خطای سناریوهای مختلف دو مدل درخت تصمیم نوع چاید و جنگل تصادفی را نشان می‌دهد. هر یک از مدل‌ها می‌توانند الگوی مشخصی برای پیش‌بینی هدایت هیدرولیکی اشباع خاک در منطقه قرقری ارائه کنند.

روند تغییرات مکانی هدایت هیدرولیکی اشباع خاک

شکل ۴ روند تغییرات مکانی هدایت هیدرولیکی اشباع خاک در منطقه مورد مطالعه را نشان می‌دهد. همان‌طور که مشاهده می‌شود، با توجه به خط روند سبز، شوری از غرب منطقه (منتهی‌الیه محور Y) به سمت شرق (منتهی‌الیه محور X) ابتدا افزایش و سپس کاهش می‌یابد. بیشترین مقادیر هدایت هیدرولیکی در نقاط مرکزی منطقه مورد مطالعه مشاهده شد. خط آبی‌رنگ روند افزایشی هدایت هیدرولیکی را نشان می‌دهد. مطابق این خط هدایت هیدرولیکی از شمال (منتهی‌الیه محور X) به سمت جنوب (منتهی‌الیه محور Y) افزایش می‌یابد. بین هدایت الکتریکی خاک و هدایت هیدرولیکی همبستگی



شکل ۴. روند تغییرات هدایت هیدرولیکی اشباع خاک

جدول ۳. نتایج ضریب تبیین و خطای سناریوهای مختلف دو مدل درخت تصمصیم نوع چاید و جنگل تصادفی

سناریو	چاید		جنگل تصادفی	
	MAE	R ²	MAE	R ²
S1	۰/۰۰۲۱۲	۰/۸۶	۰/۰۰۱۸۹	۰/۹۲
S2	۰/۰۰۲۴۵	۰/۷۹	۰/۰۰۲۰۳	۰/۸۹
S3	۰/۰۰۲۲۵	۰/۸۱	۰/۰۰۱۹۶	۰/۹۱
S4	۰/۰۰۲۸۴	۰/۶۸	۰/۰۰۲۲۵	۰/۸۵
S5	۰/۰۰۲۵۶	۰/۷۳	۰/۰۰۲۱۵	۰/۸۸
S6	۰/۰۰۲۷۸	۰/۶۹	۰/۰۰۲۶۳	۰/۸۱
S7	۰/۰۰۳۱۴	۰/۶۴	۰/۰۰۲۴۶	۰/۸۴
S8	۰/۰۰۲۶۴	۰/۷۵	۰/۰۰۲۱۱	۰/۸۷
S9	۰/۰۰۳۱۵	۰/۶۴	۰/۰۰۲۷۸	۰/۷۶
S10	۰/۰۰۲۹۳	۰/۶۶	۰/۰۰۳۰۴	۰/۷۱
S11	۰/۰۰۲۳۲	۰/۸۱	۰/۰۰۲۲۳	۰/۸۶
S12	۰/۰۰۱۹۶	۰/۸۷	۰/۰۰۱۷۵	۰/۹۴
S13	۰/۰۰۱۶۳	۰/۹۶	۰/۰۰۱۶۳	۰/۹۵
S14	۰/۰۰۴۱۲	۰/۵۷	۰/۰۰۳۷۴	۰/۶۶
S15	۰/۰۰۲۳۴	۰/۷۱	۰/۰۰۲۱۴	۰/۹
S16	۰/۰۰۲۳۸	۰/۷	۰/۰۰۱۴۷	۰/۹۸
S17	۰/۰۰۲۲۱	۰/۷۴	۰/۰۰۱۹۸	۰/۹۲
S18	۰/۰۰۲۴۵	۰/۶۷	۰/۰۰۲۰۷	۰/۸۸
S19	۰/۰۰۲۲۵	۰/۷۵	۰/۰۰۱۸۷	۰/۹۳
S20	۰/۰۰۱۸۷	۰/۹۴	۰/۰۰۱۵۶	۰/۹۷

انتقال آبخوان دشت ملکان و پیش‌بینی مقادیر تبخیر و تعرق به نتایج مشابه دست یافتند و مدل جنگل تصادفی را به عنوان مدل قوی در پیش‌بینی‌ها بیان کردند [۹ و ۱۰].

زمین‌آمار

اولین گام در استفاده از روش کریجینگ، بررسی وجود ساختار مکانی در بین داده‌ها توسط تجزیه تغییرنا است. برای این منظور، اقدام به ترسیم تغییرنا با استفاده از داده‌های اندازه‌گیری شده شد. نسبت اثر قطعه‌ای به آستانه را می‌توان برای ارزیابی ساختار مکانی داده‌ها مورد بررسی قرار داد. هنگامی که این نسبت کمتر از ۰/۲۵ باشد، متغیر مورد نظر دارای ساختار مکانی قوی است، بین ۰/۲۵-۰/۷۵ ساختار مکانی میانگین بوده و هنگامی که بزرگ‌تر از

مطابق جدول ۳ مشاهده می‌شود برای مدل چاید سناریو S13 با کمترین MAE (۰/۰۰۱۶۳) و بیشترین R² (۰/۹۶) و با ترکیب الگوی بافت خاک، EC، درصد آهک و درصد کربن آلی خاک بهترین ترکیب جهت پیش‌بینی هدایت هیدرولیکی اشباع خاک در نظر گرفته شد. در مدل جنگل تصادفی سناریو S16 با مقادیر MAE و R به ترتیب برابر ۰/۰۰۱۴۷ و ۰/۹۸ با ترکیب بافت خاک، درصد آهک و درصد ماده آلی خاک به عنوان الگوی برتر انتخاب شد. به طور کلی با توجه به مقادیر به دست آمده MAE و R مدل جنگل تصادفی نسبت به مدل چاید توانایی بیشتری برای پیش‌بینی هدایت هیدرولیکی اشباع خاک با استفاده از داده‌های زود یافت خاک دارد. محققان در تحقیق درباره پیش‌بینی مقادیر قابلیت

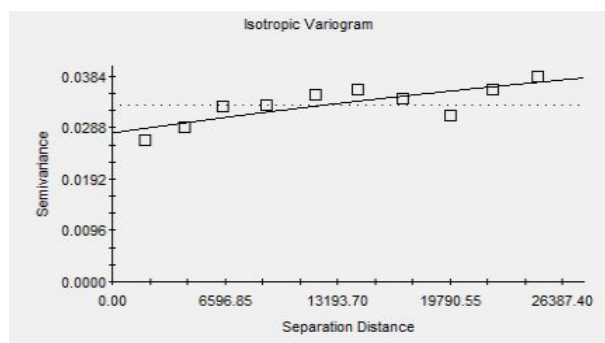
است. هر چه این معیار به یک نزدیک تر باشد، نشان دهنده برآورد خوب است. برای دستیابی به اطلاعات بیشتر، نمودار پراکنش نقاط و بهترین خط برازنده به این نقاط برای هر سه روش رسم شد (شکل ۷). با توجه به نمودارها پراکنش ارتباط مستقیم خطی بین داده‌های اندازه‌گیری شده و پیش‌بینی شده مشاهده می‌شود. بیشتر نقاط حول خط رگرسیونی و نزدیک به آن هستند. نقاطی که زیر این خط مرجع قرار گرفته‌اند، نشان دهنده این هستند که روش مقدار آن‌ها را کمتر از مقدار واقعی تخمین زده است و نقاط بالاتر از این خط نشان می‌دهند روش مقدار آن‌ها را بیشتر تخمین زده است. همان‌طور که شکل ۷ نشان می‌دهد، همه روش‌ها به خوبی هدایت هیدرولیکی اشباع خاک را پیش‌بینی کرده‌اند. با توجه به بالاتر بودن مقادیر ضریب تبیین R^2 از بین سه روش استفاده شده روش جنگل تصادفی با R^2 برابر ۰/۹۸ توانسته است از بین سه روش با دقت بیشتری مقادیر هدایت هیدرولیکی را پیش‌بینی کند.

۰/۷۵ باشد، ساختار مکانی ضعیف است [۳۱]. نتایج نشان داد مناسب‌ترین مدل برای برازش روی تغییرنما، مدل نمایی است. شکل ۵ واریوگرام ترسیم‌شده را نشان می‌دهد. مدل برازش داده‌شده نشان داد نسبت مؤلفه ساختاردار نیم‌تغییرنما به سقف $(C/(C_0+C))$ برابر ۰/۵ و نسبت اثر قطعه‌ای به حد آستانه (Nugget/sill) برابر ۴۹/۹۱ درصد به دست آمد که نشان می‌دهد وابستگی مکانی هدایت هیدرولیکی اشباع خاک در این منطقه در حد متوسطی بوده و از پیوستگی بینابینی برخوردار است.

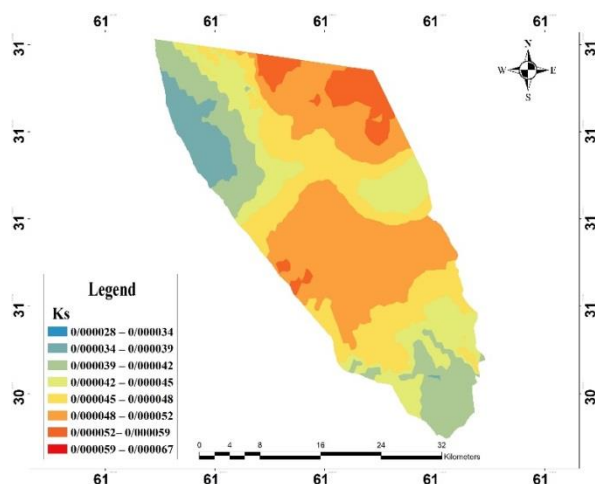
شکل ۶ توزیع مکانی هدایت هیدرولیکی اشباع خاک در منطقه مورد مطالعه را نشان می‌دهد. همان‌طور که مشاهده می‌شود، مقادیر هدایت هیدرولیکی اشباع خاک در مناطق مرکزی و شمال شرقی بیشتر است.

مقایسه روش زمین‌آمار و روش‌های چابد و جنگل تصادفی

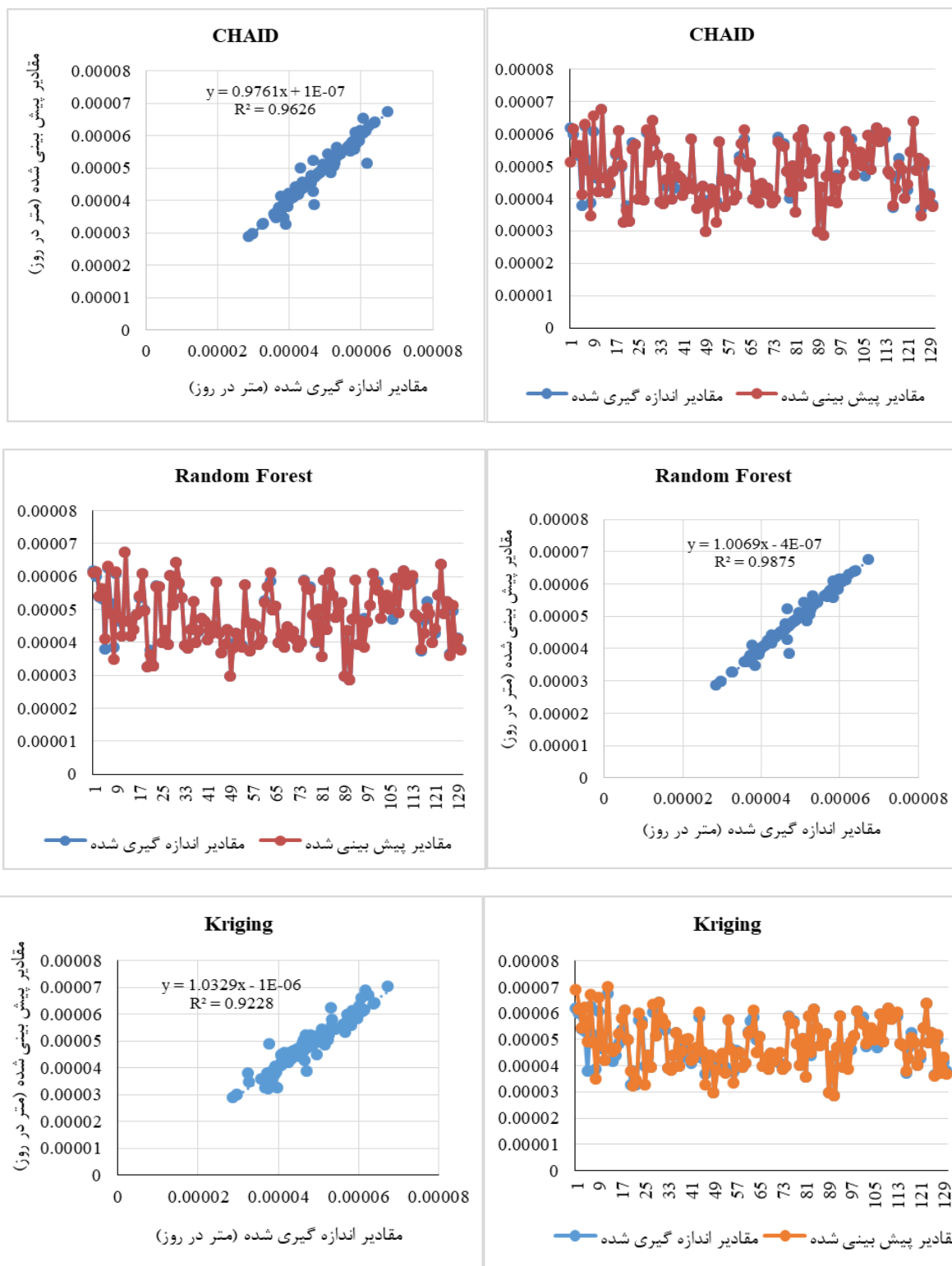
ضریب تبیین مهم‌ترین معیاری است که به کمک آن می‌توان رابطه بین دو متغیر را توضیح داد و مقدار آن بین صفر و یک



شکل ۵. تغییرنمای نمایی برازش داده‌شده



شکل ۶. پهنه‌بندی هدایت هیدرولیکی اشباع خاک در منطقه مورد مطالعه



شکل ۷. مقادیر اندازه‌گیری‌شده و پیش‌بینی‌شده هدایت هیدرولیکی اشباع خاک با استفاده از مدل چاید و جنگل تصادفی

نتیجه‌گیری

هدایت هیدرولیکی اشباع به عنوان یکی از عوامل هیدرولوژیکی از خصوصیات مهم خاک است، ولی اندازه‌گیری مستقیم آن پرهزینه و زمان‌بر است. در این

راستا استفاده از مدل‌های هوش مصنوعی برای برآورد هدایت هیدرولیکی اشباع خاک به وسیله خصوصیات زودیافت خاک مورد استفاده قرار گرفته‌اند. در تحقیق حاضر از روش‌های جنگل تصادفی و مدل چاید برای

- [7]. Rodriguez V, Ghimire B, Rogan J, Chica Olmo M, Rigol-Sánchez J.P. An assessment of the effectiveness of a Random Forest classifier for land-cover classification. *ISPRS Journal of Photogram Remote Sens.* 2012; 67: 9 -104
- [8]. Breiman L. Application and analysis of random forests and machine learning. *Journal of Water Management.* 2001; 15(1): 5-32.
- [9]. Norouzi Ghoshbalagh H, Nadiri A, Asghari Moghaddam A, Qarahkhani M. Comparison of the efficiency of artificial neural networks, fuzzy logic and random forest in estimating the aquifer transfer capability of Malekan plain. *Echo Hydrology.* 2018;5(3): 739-751 [Persian].
- [10]. Siasar H, Honar T. The application of support vector machine, chaid and random forest models in estimating daily reference transpiration evaporation in the north of Sistan and Baluchistan province. *Iran Irrigation and Drainage.* 2018 ;2(13):378-388. [Persian].
- [11]. Watt M, Palmer S. Use of regression kriging to develop a Carbon: Nitrogen ratio surface for New Zealand. *Geoderma.* 2010;183:49-57.
- [12]. Piri H, Mobaraki M, Siasar P. Temporal and spatial modeling of underground water level in Bushehr plain using artificial intelligence and geostatistics. *Watershed Management.* 2023;13(26):58-68. [Persian].
- [13]. Piri H, Mobaraki M. Comparison of Artificial Intelligence and Geostatistical Methods in Soil Surface Salinity Prediction (Case study: Ghorghori of Hirmand city). *Environ. Water Eng.* 2022;8(3): 551-537. [Persian].
- [14]. Rezaei M, Devatgar N, Tajdari Kh, Abolpur b. Investigating the spatial changes of some quality indicators of underground waters in Gilan province using geostatistics. *water and soil.* 2018;24:932-941. [Persian].
- [15]. Jahantigh M, Jahantigh M. Study effect of flood productivity on vegetation changes using field work and Landsat satellite images (Case study: Shandak of Sistan region), *Journal of RS & GIS for Natural Resources.* 2019;10(4), 57-73. [Persian].
- [16]. Gee G. W. Particle-size analysis. In: Warren, A.D. Eds. *Methods of Soil Analysis. Part 4. Physical Methods.* Soil Sci. Soc. Am. Inc. 2002; 5:255-295.
- [17]. Najibzadeh N, Ghaderi K, Ahmadi M.M. Utilization of support vector regression methods and artificial neural network in runoff precipitation modeling (Case study: Saffarud Dam catchment). *Irrigation and Drainage of Iran.* 2009;6(13):1709-1720 [Persian].

پیش‌بینی هدایت هیدرولیکی استفاده شد. همچنین روش زمین‌آمار برای بررسی همبستگی مکانی داده‌ها و پهنه‌بندی آن به کار رفت. ترکیب‌های مختلفی از بافت خاک، درصد کربن آلی خاک، درصد آهک، EC و pH به عنوان ورودی مدل‌های چاید و جنگل تصادفی در نظر گرفته شد. نتایج نشان داد هر سه مدل توانایی زیادی برای پیش‌بینی هدایت هیدرولیکی اشباع خاک دارند، اما مدل جنگل تصادفی با داشتن بیشترین R^2 (۰/۹۸) پیش‌بینی بهتری از هدایت هیدرولیکی اشباع داشت.

سپاسگزاری

این پژوهش با حمایت مالی دانشگاه زابل کد پژوهانه IR-UOZ-GR-1837 انجام شده است.

منابع

- [1]. Norouzian Azizi Z, Qajarspanlou M, Emadi S.M, Sadegh-zadeh F. Evaluation of regression models and artificial neural network in estimating hydraulic conductivity of soil saturation in Mazandaran. *Soil Research,* 2016; 31(1):76-88. [Persian].
- [2]. Bouma J. Using soil survey data for quantitative land evaluation. *Advances Soil Science.* 1989;9: 177-213.
- [3]. Wosten J.H.M, Pachepsky Ya.A, Rawls W.J. Pedotransfer functions: bridging the gap between available basic soil data and missing soil hydraulic characteristics. *J. Hydrol.* 2001;251: 123-150.
- [4]. Rezaei Arshad R, Sayad G, Mazloum M, Sharfa M, Jafarnjadi A. Comparison of artificial neural network and regression methods for predicting saturated hydraulic conductivity of soils in Khuzestan province. *Agricultural Sciences and Techniques and Natural Resources,* 2012; 16(60):107-117. [Persian].
- [5]. Sobieraj J.A, Elsenbeer H, Veressy R.A. Pedotransfer functions for estimating saturated hydraulic conductivity implications for modeling stormflow generation. *J. Hydrol.* 2001; 251: 202-220.
- [6]. Bijan A, Piri H, Tabatabai S.M, Piri J. Comparison and application of artificial neural network, support vector machine and decision tree in predicting saturated hydraulic conductivity of soil (case study: Hirmand city). *Watershed Management,* 2022;25:74-85. [Persian].

- [18]. Nelson D. W, Sommer L. E. Total carbon, organic carbon, and organic matter. *Soc. Agron., Madison*. 1982; pp. 539–579.
- [19]. Rhoads J.D. Cation exchange capacity, In; A.C. Page (ed) *Methods of soil Analysis, part 2, Am. Soc. Agron., 1986;9: 149-158.*
- [20]. Ramaswami M, Bhaskaran R. A. CHAID based performance prediction model in educational data mining. *arXiv preprint arXiv:2010;1002.1144.*
- [21]. Kisi O, Onur G, Semih D, Zounemat-Kermani M. Daily pan evaporation modeling using chi-squared automatic interaction detector, neural networks, classification and regression tree. *Computers and Electronics in Agriculture*. 2016; 122 (112-117)
- [22]. Quinlan JR. Induction of decision trees. *Journal of Machine Learning*. 1986; 1(1): 81-106.
- [23]. Soleimani K, Habibnejad M, Abkar A, Bani Asadi M. Analysis of depth, surface and continuity curves using geostatistical methods in arid and semi-arid rainfall areas (D.A.D) Case study: Sirizjan salt pan). *Desert Magazine*. 2006 ;11(1):32-41. [Persian]
- [24]. Breiman L. Application and analysis of random forests and machine learning. *Journal of Water Management*. 2001; 15(1): 5-32
- [25]. Hassani Pak A. *Geostatistics (Geostatistics)*. University of Tehran Press. 2007. p 325.
- [26]. Gholami Sh, Hosseini S.M, Mohammadi J, Mahini A.S. Spatial variability of soil macrofauna biomass and soil properties in riparian forest of Karkhe river. *Journal of Water and Soil*. 2010; 25(2):248-257 [Persian].
- [27]. Foroughifar H, Jafarzadah A.A, Torabi Gelsefidi H, Aliasgharzadah N, Toomanian N, Davatgar N. Spatial variations of surface soil physical and chemical properties on different landforms of Tabriz plain. *Journal of Soil and Water Science*. 2010;21(3):1-21 [Persian].
- [28]. Momtaz H R, Jafarzadah AA, Torabi H, Oustan Sh, Samadi A, Davatgar N, et al. An assessment of the variation in soil properties within and between Landforms in Amol region. *Iran. Geoderma*. 2009;149: 10-18.
- [29]. Ahmadi A, Palizwanzand p. Palizwanzand H. Estimation of hydraulic conductivity of soil saturation using gene expression programming and ridge regression (Case study in East Azarbaijan province). *Iranian Soil and Water Research*. 2017;48 (5):1087-1095 [Persian].
- [30]. Indirect estimation of near-saturated hydraulic conductivity from readily available soil information. *Geoderma*. 2002;108, 1-17.
- [31]. Shi J, Wang H, Xu J, Wu J, Liu X, Zhu H, Yu C. Spatial distribution of heavy metals in soils: a case study of Changxing. *China. Environ Geol*. 2007 ; (52):1-10.