دانشگاه علوم کشاورزی و منابع طبیعی گرگان
نشریه پژوهش‌های حفاظت آب و خاک
جلد بیستم، شماره پنجم، ۱۳۹۲
http://jwsc.gau.ac.ir

مقایسه روش‌های زمین‌آماری با روش غیرپارامتریک *k*– *نزدیک‌ترین همسایه* برای برآورد هدایت هیدرولیکی اشباع خاک

***وحیدرضا جلالی[۱]، عباس خاشعی‌سیوکی[۲] و مهدی همایی[۳]**

[۱]استادیار گروه خاکشناسی، دانشگاه شهید باهنر کرمان، [۲]استادیار گروه مهندسی آب، دانشگاه بیرجند،
[۳]استاد گروه خاکشناسی، دانشگاه تربیت مدرس

تاریخ دریافت: ۹۱/٤/۳۱ ؛ تاریخ پذیرش: ۹۱/۸/۱۵

چکیده

آگاهی از هدایت هیدرولیکی اشباع خاک در مدیریت آبیاری و مدل‌سازی هیدورلوژیکی، ضروری می‌باشد، ولی در بیش‌تر موارد به‌علت محدودیت‌های عملی و یا هزینه‌ای، اندازه‌گیری آن با دشواری همراه است. در این پژوهش مدل‌های مختلف زمین‌آماری با نوعی از الگوریتم‌های غیرپارامتریک از نوع یادگیرنده‌های تنبل موسوم به *k*– *نزدیک‌ترین همسایه*، برای تخمین هدایت هیدرولیکی اشباع خاک از روی داده‌های سهل‌الوصول خاک، مورد مقایسه قرار گرفت. در این پژوهش ۱۵۱ نمونه از خاک‌های زراعی اطراف بجنورد، جمع‌آوری شده و متغیرهای کمکی شامل فراوانی ذرات، جرم مخصوص حقیقی و ظاهری همچنین هدایت الکتریکی عصاره اشباع خاک (EC_e)، درصد مواد آلی خاک (OM)، رطوبت اشباع خاک (θ_s)، pH و میزان مواد خنثی‌شونده آن (TNV) برای برآورد هدایت هیدرولیکی اشباع به‌کار گرفته شد. استفاده از آماره‌های ضریب هبستگی پیرسون (r)، خطای ماکزیمم (ME)، ریشه میانگین مربعات خطا (RMSE)، کارآیی مدل (EF) و ضریب جرم باقی‌مانده (CRM) نشان داد که از نظر دقت برآورد، روش غیرپارامتریک *k*– *نزدیک‌ترین همسایه* (با داشتن آماره‌های r=۰/۷٦، EF=۰/٦٥٥، RMSE=٤۲/۸۷، ME=۲٦/۸۹ و CRM= –۰/۱۱) نسبت به سایر روش‌ها از دقت قابل‌قبولی برخوردار بوده و پس از آن روش‌های کوکریجینگ و کریجینگ ساده، واجد بیش‌ترین دقت در تخمین مقادیر

* مسئول مکاتبه: v.jalali@uk.ac.ir

هدایت هیدرولیکی اشباع خاک بوده‌اند. بر این اساس، می‌توان نتیجه‌گیری کرد که استفاده از تکنیک k- *نزدیک‌ترین همسایه* به‌عنوان روشی جایگزین برای اشتقاق توابع انتقالی خاک، به‌ویژه هنگامی‌که فراهمی داده‌های جدید، نیاز به اشتقاق دوباره این توابع را الزام‌آور می‌کند، می‌تواند به‌کار رود.

واژه‌های کلیدی: تکنیک k- *نزدیک‌ترین همسایه*، روش‌های زمین‌آماری، هدایت هیدرولیکی اشباع خاک

مقدمه

یکی از مشکلات مهمی که کاربرد یافته‌های مربوط به ویژگی‌های هیدرولیکی خاک را در مورد پیش‌بینی و ارزیابی فرآیندهای موجود در محیط متخلخل خاک متأثر می‌سازد، تغییرپذیری مکانی[1] توابع هیدرولیکی خاک است. این تغییرپذیری چنان گسترده است که می‌توان آن را در مقیاس‌هایی از یک مترمربع تا بیش از ۱۵۰ هکتار مشاهده کرد (بیگار و نیلسن، ۱۹۷٦). نکته قابل‌توجه در این باره آن است که این تغییرپذیری هم در مقدار توابع و هم در الگو و ساختار آن قابل مشاهده است. مالانتس و همکاران (۱۹۹٦) شدت تغییرپذیری ویژگی‌های هیدرولیکی خاک را در ۳ دسته طبقه‌بندی نموده‌اند که عبارتند از خصوصیات به‌شدت تغییرپذیر مانند هدایت هیدرولیکی اشباع (k_s)، خصوصیات با تغییرات متوسط مانند میزان رطوبت باقی‌مانده (θ_r) و خصوصیات با تغییرپذیری اندک مانند مقدار رطوبت اشباع خاک (θ_s). از آن‌جایی‌که دانش فیزیک خاک و هیدرولوژی محیط‌های متخلخل، همواره با پدیده‌هایی که تغییرپذیری مکانی و زمانی بالایی دارند مواجه می‌باشند، بنابراین برون‌یابی خطی نتایج به‌دست آمده از نمونه‌گیری نقطه‌ای و پراکنده آزمایشگاهی و یا کرت‌های مزرعه‌ای، هنگامی‌که به مقیاس‌های بزرگ‌تری تعمیم یابند، دچار نبود قطعیت زیادی می‌شوند.

راه‌حل‌هایی متفاوت برای حل این مشکل توسط پژوهش‌گران مختلف ارایه شده است که از آن جمله می‌توان به همبستگی تجربی (correlation) بین متغیرهای خاکی مانند توابع انتقالی (بوما، ۱۹۸۹)، استفاده از نظریه مقیاس محیط‌های مشابه (similar media scaling theory) (میلر و میلر، ۱۹۵٦)، استفاده از تئوری متغیرهای ناحیه‌ای و روش‌های زمین‌آماری (Geostatistical techniques) (جانگ و همکاران، ۲۰۰٦) اشاره کرد. با این وجود، هر یک از این روش‌ها دارای محدودیت‌های ویژه‌ای بوده که کاربرد آن‌ها را منحصر به شرایطی خاص و مقیاسی معین می‌کند.

1- Spatial Variability

امروزه کاربرد روش‌های زمین‌آماری در مطالعه ویژگی‌های هیدرولیکی و فیزیکی خاک در مطالعات زیادی مورد استفاده قرار گرفته است (سوبیراج و همکاران، ۲۰۰٤). ولی از آن‌جا که این روش، تنها یک روش آماری مرتبه دوم[1] است، در حضور مقادیر کم و زیاد در سری داده‌ها، تنها قادر به بررسی تغییرپذیری مکانی متغیر به‌صورت ضعیف می‌باشد (کارواچنکو و همکاران، ۲۰۰۰). به‌علت وجود رابطه غیرخطی بین قطر منافذ و شدت جریان، داده‌های مربوط به ویژگی‌های هیدرولیکی خاک معمولاً از توزیعی نرمال برخوردار نیستند. بنابراین توزیع‌هایی با چولگی زیاد و ضریب تغییرات بیش از ٤۰۰ درصد، زیاد دور از انتظار نیست (لی و همکاران، ۱۹۸۵). بر این اساس، شرط لازم ایستا بودن (stationarity) داده‌ها و نیز لزوم وجود توزیع نرمال و یا حداقل شبه‌نرمال (quasi-Gaussian) داده‌ها نیز تردیدهای جدی در استفاده مستقیم از روش‌های زمین‌آماری برای مطالعه داده‌های هیدرولیکی با تغییرپذیری بسیار زیاد و نیز توزیع غیرنرمال به وجود آورده است (پرینگل و همکاران، ۲۰۰۷). بنابراین به‌دلیل نبود اطلاعات و دانش کافی در زمینه شیوه نشان دادن تغییرپذیری متغیرهای خاکی در مقیاس‌های مختلف، این موضوع به چالشی مهم در فیزیک خاک و هیدرولوژی تبدیل شده است.

آگاهی از هدایت آبی اشباع خاک برای درک و مدل‌سازی بسیاری از فرآیندهای فیزیکی خاک ضروری است. تعیین مقدار رواناب سطحی و نفوذ به خاک، ماندگاری موقتی آب در محیط ریشه، نرخ انتقال املاح و بسیاری دیگر از فرایندهای کشاورزی و زیست‌محیطی وابسته به‌میزان هدایت هیدرولیکی اشباع خاک (K_s) می‌باشد (زلک و سی، ۲۰۰۵). با وجود پیشرفت‌های تکنیکی و بهبود ابزارآلات به‌کار رفته در اندازه‌گیری مستقیم این ویژگی خاک، این روش‌ها هم‌چنان زمان‌بر و همراه با خطا می‌باشند. بنابراین پژوهش‌گران برای حل این مشکل، روش‌های غیرمستقیم را مورد توجه قرار داده‌اند تا به‌وسیله آن بتوانند در نقاط بدون داده، تخمینی بهینه از متغیر موردنظر داشته باشند. اشتقاق توابع انتقالی خاک و استفاده از روش‌های زمین‌آماری از جمله این روش‌هاست. در توابع انتقالی، از روی ویژگی‌های زودیافت خاک، ویژگی‌های دیریافت آن را برآورد می‌کنند (بوما، ۱۹۸۹). تکنیک‌های رگرسیونی و اخیراً، شبکه‌های عصبی مصنوعی دو روش معمول در توسعه توابع انتقالی خاک می‌باشند. اسخاپ و همکاران (۲۰۰۱) با استفاده از نرم‌افزار ***Rosetta*** (توابعی انتقالی براساس شبکه‌های عصبی مصنوعی)، اقدام به تخمین و برآورد ویژگی‌های هیدرولیکی خاک نمودند.

1- Second Order Moment

وجه تشابه بین بیش‌تر توابع انتقالی موجود، در اشتقاق آن‌ها بر مبنای رویکرد پارامتریک می‌باشد. به این معنی که همه این توابع متشکل از پارامترهایی هستند که از برازش یک‌سری توابع معین بر داده‌ها به‌دست آمده‌اند که این روی‌کرد، خود کاستی‌هایی به همراه دارد.

تعیین رابطه صحیح و حصول اطمینان از یکنواختی توزیع تابع احتمال خطا در بین داده‌ها، معمولاً کار ساده‌ای نیست. هم‌چنین هنگامی‌که بانک داده از تعدادی اندک تشکیل شده باشد، تخمین‌های شکل گرفته براساس روی‌کرد نام برده، بسیار ناپایدار خواهد بود. و از طرفی، در مواردی که داده‌های جدید (در مقیاس زمانی و مکانی متفاوت با داده‌های موجود در بانک مرجع) مهیا گردد، بازنگری کلی در رابطه‌های قبلی و توسعه دوباره آن‌ها، الزام‌آور خواهد شد. به همین دلیل کاربران به‌راحتی قادر به اضافه نمودن داده‌های محلی خود برای بهبود تخمین این توابع نیستند (نمس و همکاران، ۲۰۰۸).

در علم زمین‌آمار اختلاف پدیده‌ها با توجه به مکان و زمان بررسی می‌شود. با استفاده از تکنیک زمین‌آمار می‌توان سطحی پیوسته از خصوصیات آماری نقاط معلوم را به‌وجود آورد. با توجه به این‌که تغییرات بسیاری از پدیده‌های طبیعی مانند ویژگی‌های هیدرولیکی خاک، تابع زمان و مکان است، استفاده از آمار کلاسیک که در آن اختلاف دو نقطه در فضا مستقل از فاصله مکانی و زمانی در نظر گرفته می‌شود، نمی‌تواند به‌صورت مؤثری تغییرات این پدیده‌ها را تفسیر نماید (دوچ، ۲۰۰۲). روش‌های زمین‌آماری به‌دلیل در نظر گرفتن موقعیت و آرایش داده‌ها و هم‌چنین همبستگی مکانی آن‌ها در برخی موارد دقت مناسبی دارند ولی کاربرد روش‌های زمین‌آماری مانند کریجینگ هنگامی‌که تعداد اطلاعات اندازه‌گیری شده محدود، غیرهمگن و دارای نبود اطمینان و نبود قطعیت باشند، محدود می‌گردد (وبستر و اُلیویر، ۲۰۰۷).

استفاده از تکنیک‌های غیرپارامتریک می‌تواند به‌عنوان روی‌کردی جایگزین، برای این‌چنین تخمین‌هایی به‌کار گرفته شود. این تکنیک‌ها، به‌جای برازش دادن یک‌سری توابع معین بر داده‌ها، براساس تشخیص الگو و استفاده از اصل تشابهات بنا نهاده شده‌اند. به‌عنوان نمونه اسخاپ و همکاران (۲۰۰۱) نرم‌افزار ***Rosetta*** را براساس روی‌کرد پارامتریک و با استفاده از شبکه‌های عصبی مصنوعی ابداع نموده بودند، در پژوهشی نوین (۲۰۰۹) با استفاده از همان پایگاه داده‌ای که در اشتقاق توابع هیدرولیکی خاک به‌کار برده بودند، به این نتیجه رسیدند که استفاده از تکنیک‌های غیرپارامتریک کارایی چشم‌گیری در بهبود تخمین‌های صورت‌گرفته خواهد داشت (تاوارسکی و همکاران، ۲۰۰۹). به همین ترتیب نمس و همکاران (۲۰۰۹) توابع انتقالی که توسط راولز و همکاران در سال ۱۹۸۲ به

روش رگرسیون خطی اشتقاق یافته بودند را بررسی نمودند و دریافتند که این توابع انتقالی از دقت کافی برای استفاده آن‌ها در مقیاس ایالات متحده آمریکا برخوردار نبوده و با ارزیابی روش غیرپارامتریک *k- نزدیک‌ترین همسایه*، بیان نمودند که روش نام برده از توانایی بالاتری برای تخمین توابع هیدرولیکی در مقیاس کل ایالات متحده آمریکا برخوردار است.

پژوهش نمس و همکاران (۱۹۹۹) در رابطه با تفسیر توزیع اندازه ذرات خاک با کمک تکنیک *k- نزدیک‌ترین همسایه*، در واقع یکی از اولین موارد استفاده این تکنیک در علوم خاک بوده است. پس از اثبات توانایی روش نام برده، نمس و همکاران (۲۰۰۹) با استفاده از تکنیک *k- نزدیک‌ترین همسایه*، به تخمین ویژگی‌های هیدرولیکی خاک پرداختند و بیان نمودند که بر خلاف دقت یکسان تکنیک نام برده و روش شبکه‌های عصبی مصنوعی در اشتقاق و تخمین توابع هیدرولیکی، قابلیت روش *k- نزدیک‌ترین همسایه* برای وارد نمودن داده‌های محلی، برتری نسبی برای این روش ایجاد می‌نماید.

سیگال و همکاران (۲۰۰۸) نیز به مطالعه حرکت و انتقال آب و املاح در مقیاس ناهمگون مزرعه‌ای پرداختند. ایشان در پژوهش خود از تکنینک *k- نزدیک‌ترین همسایه* به‌عنوان الگوریتم میان‌یابی نام‌برده و به‌طور موفقیت‌آمیزی توانستند با استفاده از این روی‌کرد، درک عمیقی از تغییرپذیری ویژگی‌های هیدرولیکی خاک در مقیاس مزرعه‌ای به‌دست آورند.

بر خلاف توابع انتقالی کلاسیک، تکنیک *k- نزدیک‌ترین همسایه* از هیچ تابع ریاضیاتی از پیش تعریف‌شده‌ای برای تخمین متغیرهای مختلف استفاده نمی‌نماید. در این روی‌کرد، یک بانک داده مرجع[1]- همانند بانک داده‌ای که در آموزش و توسعه توابع انتقالی کلاسیک به‌کار می‌رود- برای یافتن نزدیک‌ترین (مشابه‌ترین) خاک به خاک هدف، مورد جستجو واقع می‌شود. توضیح کامل این روش در منابع مختلف ذکر شده است (جلالی و همایی، ۲۰۱۱).

بر این اساس هدف اصلی این پژوهش، مقایسه کارایی نسبی روش غیرپارامتریک K-nn با روش‌های معمول زمین‌آماری در تخمین ضریب هدایت هیدرولیکی اشباع خاک براساس سایر متغیرهای در دسترس بوده است.

1- ‘Reference’ Data Set

مواد و روش‌ها

منطقه مورد مطالعه، دشت دامنه‌ای قره‌میدان واقع در ۷۰ کیلومتری شمال‌غرب بجنورد بود. وسعت منطقه بیش از ۳۰۰ هکتار و شیب عمومی آن حدود ۱۵ درصد می‌باشد. در آغاز پژوهش، با استفاده از نرم‌افزار ArcGIS و دستگاه GPS، کل منطقه به شبکه‌هایی با طول مساوی ۱۵۰ متر تقسیم‌بندی شد. از بخش‌های نام برده نیز از عمق ۲۵-۰ سانتی‌متری نمونه‌برداری خاک انجام شد و تعداد ۱۵۱ نمونه خاک انتخاب گردید.

در این پژوهش از دستگاه نفوذسنج گلف برای تعیین ضریب آب‌گذری اشباع خاک استفاده گردید. روش نفوذسنج گلف (رینولدز و الریک، ۱۹۸۷) یکی از روش‌های اندازه‌گیری نفوذپذیری تحت بار ثابت می‌باشد. در این روش دبی ثابت آب خروجی از چاهک به خاک اطراف تحت بار آبی ثابت اندازه‌گیری می‌شود. جرم ویژه ظاهری نمونه‌ها به روش کلوخه و جرم ویژه حقیقی از طریق پیکنومتر (کارتر، ۱۹۹۳) تعیین شد. نمونه‌ها در مجاورت هوای آزاد خشک و از الک ۲ میلی‌متری عبور داده شدند. فراوانی نسبی اندازه ذرات به روش هیدرومتری اندازه‌گیری و کلاس بافتی خاک‌ها تعیین گردید (کلوت، ۱۹۸٦).

همچنین سایر ویژگی‌های خاک شامل، هدایت الکتریکی عصاره اشباع خاک (EC_e) و pH (پیج و همکاران، ۱۹۸۲)، درصد مواد آلی خاک (OM) (والکلی و بلک، ۱۹۳٤)، رطوبت اشباع خاک (θ_s) (کارتر، ۱۹۹۳) و میزان مواد خنثی‌شونده آن نیز (TNV) (کلوت، ۱۹۸٦) تعیین گردید.

در مرحله بعد، با استفاده از رابطه لال و شارما (۱۹۹٦)، تخمینی اولیه از تعداد k نمونه برای وارد نمودن در محاسبه‌ها به‌عمل آمد و سپس با استفاده از تکنیک ارزیابی تقاطعی، تعداد دقیق k نمونه خاک محاسبه گردید.

تکنیک Cross Validation روشی آماری است که در واقع کیفیت تخمین‌های یک مدل را براساس تعداد و نوع داده‌های ورودی مشخص می‌کند. نام دیگر این تکنیک، تخمین چرخشی (rotation estimation) می‌باشد زیرا با جای‌گذاری n تعداد از داده‌های ورودی، به تخمین متغیر مجهول پرداخته و میزان اختلاف داده تخمینی و مشاهده‌ای را ثبت می‌کند، در گردش بعدی میزان اختلاف بین داده تخمینی و مشاهده‌ای را برای n+۱ تعداد از داده‌های ورودی محاسبه می‌کند. این چرخش تا زمانی ادامه می‌یابد تا سرانجام، ترکیب موردنظر که دارای کم‌ترین خطا بین مقادیر تخمینی و مشاهده‌ای است، به‌دست آید.

پس از تعیین تعداد k- *نزدیک‌ترین همسایه* برای ورود به محاسبه‌ها، برنامه موردنظر برای ورود ویژگی‌های هر خاک شامل مختصات جغرافیایی هر نقطه در سیستم متریک، درصد ذرات شن، سیلت و رس، درصد مواد خنثی‌شونده (TNV)، هدایت الکتریکی عصاره اشباع خاک (EC_e)، مقدار رطوبت اشباع آن (θ_s) و جرم ویژه ظاهری و حقیقی خاک‌ها، در محیط برنامه‌نویسی R اجرا گردید. فواصل اقلیدسی داده هدف با هر یک از داده‌های بانک مرجع محاسبه و ذخیره گردید. مقدار تخمینی برای هر کدام از نمونه‌های هدف براساس میانگین وزنی k تعداد از نزدیک‌ترین همسایه‌های از پیش تعیین‌شده، به‌دست آمد. پس از آن، در محیط نرم‌افزار Arc-Gis با استفاده از روش‌های مختلف میان‌یابی شامل: کوکریجینگ (Co-Kriging)، کریجینگ جامع (Universal Kriging)، کریجینگ معمولی (Ordinary Kriging)، کریجینگ ساده (Simple Kriging)، روش عکس مجذور فاصله (Inverse distance weights)، روش تابع شعاعی (Radial Basis function)، روش میان‌یابی چندجمله‌ای محلی (Local polynomial interpolation) و روش میان‌یابی چندجمله‌ای جهانی (Global polynomial interpolation) اقدام به تخمین مقادیر ضریب هدایت هیدرولیکی اشباع شد. در این قسمت از پژوهش برای استفاده از روش کوکریجینگ از متغیر جرم مخصوص ظاهری، به‌عنوان داده کمکی استفاده شد، زیرا این پارامتر بیش‌ترین همبستگی را با پارامتر هدایت هیدرولیکی داشت.

در نهایت اقدام به ارزیابی و اعتبارسنجی عملکرد مدل‌ها با استفاده از یک‌سری شاخص‌های آماری شد. یکی از شاخص‌های آماری که برای ارزیابی مدل‌ها از آن استفاده می‌شود، ضریب همبستگی پیرسون می‌باشد که توسط رابطه زیر تعریف می‌شود (رابطه ۱):

$$r=\frac{n\left(\sum_{i=1}^{n}(P_i)(O_i)\right)-\left(\sum_{i=1}^{n}P_i\right)\left(\sum_{i=1}^{n}O_i\right)}{\sqrt{n\left[\sum_{i=1}^{n}(P_i)^2-\left(\sum_{i=1}^{n}(P_i)\right)^2\right]\left[n\sum_{i=1}^{n}(O_i)^2-\left(\sum_{i=1}^{n}(O_i)\right)^2\right]}},\quad -1\le r\le 1 \qquad (1)$$

که در آن، r: ضریب همبستگی، P_i: مقدار پیش‌بینی شده برای نمونه i ام و O_i: مقدار مشاهده شده برای نمونه i ام می‌باشد. از آن‌جا که مقادیر ضریب همبستگی همواره در بازه [۱ و ۱-] قرار می‌گیرند، قضاوت از روی این ضریب ساده است و ممکن است به‌نظر برسد که ضریب همبستگی می‌تواند معیار مناسبی در ارزیابی مدل باشد. با این‌حال باید توجه داشت که ضریب همبستگی نمی‌تواند به تنهایی شاخص مناسبی برای ارزیابی مدل باشد. زیرا ممکن است در یک مدل فرضی مقادیر پیش‌بینی و

مشاهده شده دارای اختلافی فاحش باشند ولی این اشتباهات به‌گونه‌ای باشد که از یک روند یکنواخت پیروی نماید. بنابراین اگرچه ضریب همبستگی به‌خوبی نشان‌دهنده میزان هم‌آهنگی روند تغییرات مقادیر مشاهده شده نسبت به مقادیر پیش‌بینی شده می‌باشد اما گویای تطابق آن‌ها نیست (قربانی‌دشتکی و همکاران، ۲۰۰۹). شاخص‌های کمی دیگری که می‌توان در برآورد دقت مدل از آن‌ها استفاده نمود، عبارتند از آماره‌های خطای ماکزیمم (ME)[1]، ریشه میانگین مربعات خطا (RMSE)[2]، کارآیی مدل (EF)[3] و ضریب جرم باقی‌مانده (CRM)[4]. بیان ریاضی آماره‌های نام برده به‌صورت زیر است:

$$ME = \max \left| P_i - O_i \right|_{i=1}^{n} \quad (2)$$

$$RMSE = \left[\frac{\sum_{i=1}^{n} (P_i - O_i)^2}{n} \right]^{\frac{1}{2}} \quad (3)$$

$$EF = \frac{\sum_{i=1}^{n} \left(O_i - \overline{O}\right)^2 - \sum_{i=1}^{n} (P_i - O_i)^2}{\sum_{i=1}^{n} \left(O_i - \overline{O}\right)^2} \quad (4)$$

$$CRM = \frac{\sum_{i=1}^{n} O_i - \sum_{i=1}^{n} P_i}{\sum_{i=1}^{n} O_i} \quad (5)$$

که در آن‌ها، P_i: مقادیر برآورد شده، O_i: مقادیر اندازه‌گیری شده و n: تعداد نمونه است. اگر تمامی داده‌های برآورد شده و اندازه‌گیری شده یکسان باشند، نتایج این آماره‌ها به‌صورت ME=۰، RMSE=۰، EF=۱ و CRM=۰ خواهد بود.

نتایج و بحث

جدول ۱، توصیفی آماری از ویژگی‌های انتخابی خاک‌های موجود در بانک داده را نشان می‌دهد.

1- Maximum Error
2- Root Mean Square Error
3- Efficiency of Model
4- Coefficient of Residual Mass

جدول ۱- خلاصه‌ای از آماره‌های توصیفی پارامترهای خاکی به‌کار رفته برای تخمین k نزدیک‌ترین همسایه.

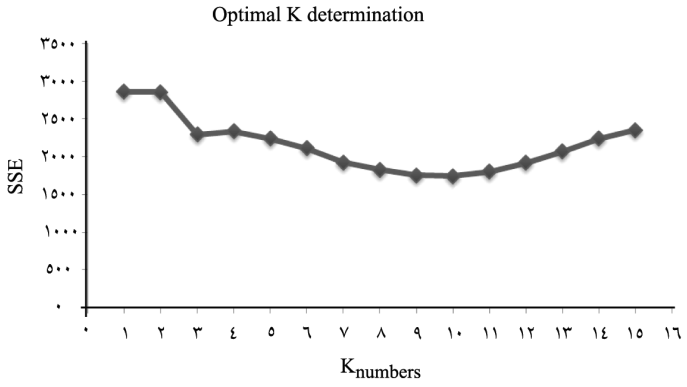
ویژگی	واحد	دامنه	کمینه	بیشینه	میانگین	انحراف معیار	ضریب تغییرات
جرم ویژه ظاهری	گرم بر سانتی‌مترمکعب	۰/٤۹	۱/۲٦	۱/۷۵	۱/۵۳	۰/۰۸٦	۵/٦۲
جرم ویژه حقیقی	گرم بر سانتی‌مترمکعب	۰/۵۹	۲/۱۹	۲/۷۸	۲/۵٦	۰/۰۸۸	۳/٤۳
شن	گرم بر گرم	۰/۷۱	۰/۰٤	۰/۷۵	۰/۲۱۰۷	۰/۰۸۳۸۲	۳۹/۷۷
سیلت	گرم بر گرم	۰/٤۹	۰/۱۳	۰/٦۲	۰/٤۸۷۰	۰/۰۵٤۰۸	۱۱/۱۰
رس	گرم بر گرم	۰/۳۲	۰/۱۲	۰/٤٤	۰/۳۰۲۳	۰/۰٤٤۸۲	۱۵
کربن آلی	درصد	۱/۷۰	۰/۲۱	۱/۹۱	۰/۸۸	۰/۲۲	۲٤/۹
مواد خنثی‌شونده	درصد	٤۲	۵/۷۵	٤۷/۷۵	۲۱/۹۹	۵/۳٦	۲٤/۳۸
شوری	دسی‌زیمنس بر متر	۳/۵٤	۰/۲۷	۳/۸۱	۱/۷٤	۰/٤۸	۲۸
رطوبت اشباع (θ_s)	متر بر مترمکعب	۰/۰۹	۰/۳۸	۰/٤٦	۰/٤۰۷	۰/۰۱۲٦	۳/۱
هدایت هیدرولیکی اشباع خاک (ks)	سانتی‌متر بر d	۱۹۳/۲۲	۱/۳۹	۱۹٤/٦۲	۱۷/۱۲	۱۹/۸۵	۱/۱٦

همان‌گونه که اشاره شد، رابطه لال و شارما (۱۹۹٦) تخمینی اولیه از مقدار k تعداد بهینه از نزدیک‌ترین همسایه‌ها از بانک مرجع به داده هدف ارایه می‌دهد. با توجه به این‌که تعداد داده‌های موجود در بانک مرجع ۱۵۱ (n>۱۰۰) عدد بود، بنابراین:

$$k = n^{۱/۲} \quad for\ n > ۱۰۰$$
$$k = ۱۵۱^{۱/۲} = ۱۲/۳$$

عدد به‌دست آمده حدود تقریبی میزان k بهینه را نشان می‌دهد، ولی برای تعیین دقیق عدد k، از تکنیک ارزیابی تقاطعی استفاده شد. شکل ۱ مقدار دقت در تکنیک ارزیابی تقاطعی برای تعیین تعداد k بهینه را براساس آماره مجموع مربعات خطا (SSE)[1] نشان می‌دهد. همان‌طورکه مشاهده می‌شود، مقدار خطا در کم‌ترین تعداد همسایگی یعنی K=۱ حداکثر بوده و با افزایش این تعداد از میزان خطا کاسته شده است. این روند تا K=۱۰ ادامه یافته ولی پس از آن دوباره میزان خطا افزایش یافته و یا به‌عبارت دیگر دقت تخمین‌ها کاهش یافته است. پس در واقع در دامنه موردنظر (۱۵-۱) تعداد K=۱۰ بهینه‌ترین تعداد همسایگی برای انجام تخمین‌ها بوده است.

1- Sum of Square Error



شکل ۱- تعیین میزان k بهینه براساس آماره مجموع مربعات خطا (SSE).

در مرحله بعد الگوریتم موردنظر در محیط برنامه R نوشته شد (ورزانی، ۲۰۰٤) تا خود برنامه به شکل هوشمند از بانک داده، نزدیک‌ترین داده‌ها را به داده هدف انتخاب نموده و پس از مرتب‌سازی آن‌ها براساس کم‌ترین فاصله اقلیدسی، ۱۰ داده نزدیک به داده هدف را جدا و با میانگین وزنی، مقدار هدایت هیدرولیکی اشباع خاک هدف را تخمین بزند. تخمین‌های زمین‌آماری نیز در محیط برنامه ArcGis انجام پذیرفت که نتایج آن در جدول ۲ و شکل ۲ نشان داده شده است.

در این پژوهش علاوه‌بر محاسبه ضریب همبستگی، از آماره‌های خطای ماکزیمم (ME)، ریشه میانگین مربعات خطا (RMSE)، کارآیی مدل (EF)، ضریب جرم باقی‌مانده (CRM) و میانگین مطلق خطا (MAE) به‌عنوان شاخص‌هایی کمی برای تعیین میزان کارایی مدل‌های مختلف استفاده شد. جدول ۲ میزان هر کدام از این آماره‌ها را نشان می‌دهد.

جدول ۲- آماره‌های محاسباتی برای تعیین میزان قابلیت تکنیک‌های مختلف در برآورد هدایت هیدرولیکی خاک.

روش میان‌یابی	r	EF	RMSE	ME	CRM	MAE	رتبه مدل
نزدیک‌ترین همسایه k	۰/۷٦	۰/٦٥٥	٤۲/۸۷	۲٦/۸۹	-۰/۱۱	٤/۲۹	۱
Co-Kriging	۰/۷۱	۰/٥۹	٤۱/٥۲	۳۰/۸۱	-۰/۰۸۱	٤/۰۲	۲
Simple kriging	۰/٥۹	۰/٤٦	٤۷/٦۲	۳٤/۰۱	-۰/۰۷	٤/٦۷	۳
Universal kriging	۰/٥۹	۰/٤٦	٤۷/۸	۳٤/۸٦	-۰/۰۸	٤/۷۲	٤
Original kriging	۰/٥۹	۰/٤٥	٤۷/۸	۳۳/۱٤	-۰/۰۸	٤/۷٥	٥
IDW	۰/٥۹	۰/٤٥	٤۸/٥۸	۳٥/٤	-۰/۱۲	۷/۱٤	٦
Radial Basis Function	۰/٥۳	۰/۳۸	٥۱/۰۸	۳۳/۱۲	-۰/۱۱	٥/۳	۷
Local polynomial	۰/٥	۰/۳٦	٥۱/۸۷	۳٥/٤	-۰/۱۱	٥/۲٦	۸
Global polynomial	۰/٤٦	۰/۳۱	٥۳/۸۲	۳٤/۸۹	-۰/۱۲۹	٥/٥۸	۹

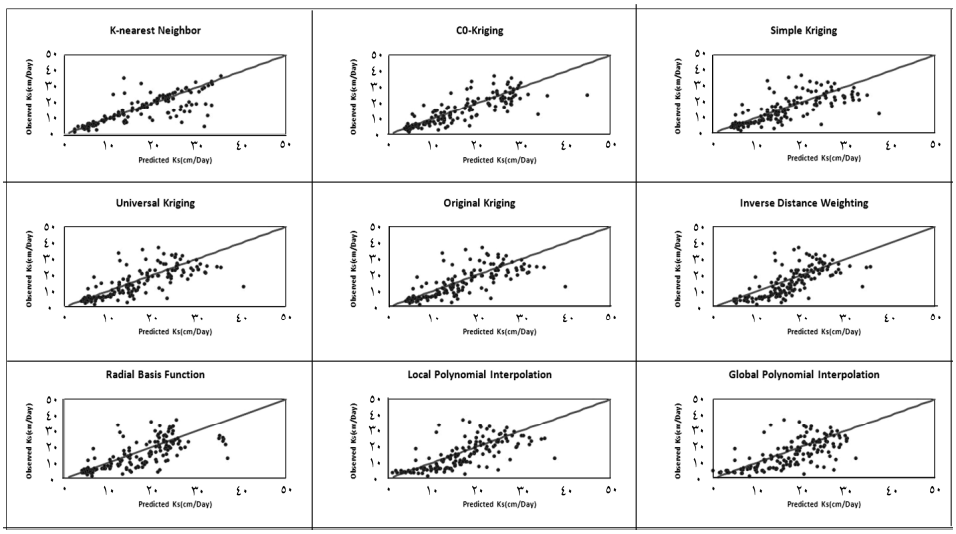
با توجه با داده‌های جدول مشاهده می‌شود که روش غیرپارامتریک *k- نزدیک‌ترین همسایه* نسبت به سایر روش‌ها از دقت قابل‌قبولی برخوردار بوده و پس از آن روش‌های کوکریجینگ و کریجینگ ساده، واجد بیش‌ترین دقت در تخمین مقادیر هدایت هیدرولیکی اشباع خاک بوده‌اند.

از روی شکل ۲ نیز می‌توان نتیجه گرفت که روند تغییرات مقادیر برآوردی توسط مدل با مقادیر اندازه‌گیری شده، در روش‌های مختلف، تا حدی هم‌آهنگ بوده ولی در روش غیرپارامتریک *k- نزدیک‌ترین همسایه*، روند داده‌های اندازه‌گیری شده و برآورد شده پراکنش کم‌تری از خط ۱:۱ داشته‌اند. یا به‌عبارت دیگر مدل نام برده توانسته است با دقت به‌نسبت خوبی مقادیر هدایت هیدرولیکی اشباع خاک را برآورد نماید (r=۰/۷۶) و به همین ترتیب در شکل ۳، نقشه پراکنش میزان هدایت هیدرولیکی اشباع اندازه‌گیری شده و برآورد شده توسط ۳ تکنیکی که توانمندی بیش‌تری در انجام تخمین‌های خود داشته‌اند، نشان داده شده است.

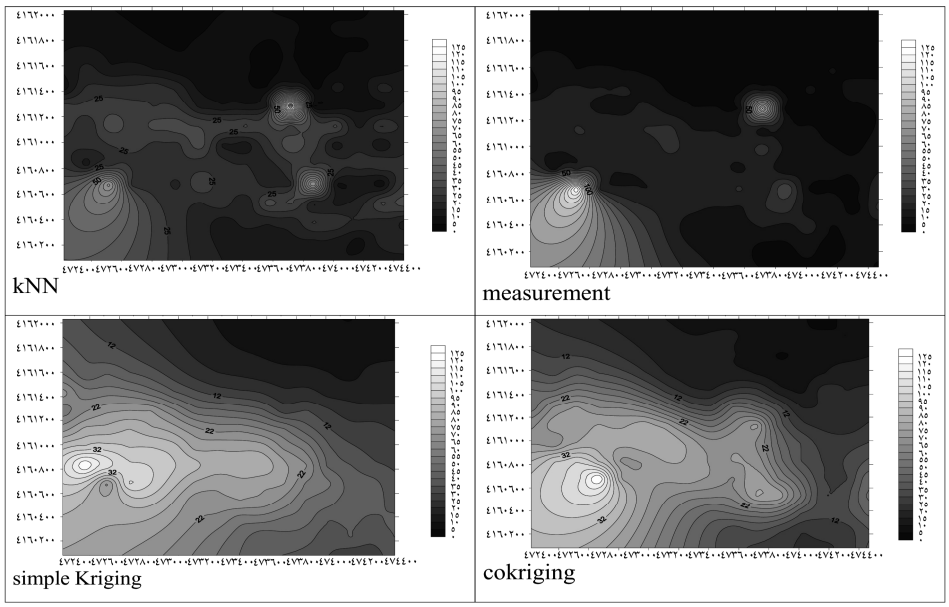
به‌طورکلی می‌توان گفت که با در نظر گرفتن داده‌های جدول ۲، تکنیک‌های مختلف ۹ گانه، توانسته‌اند تا حدی، مقادیر ضریب هدایت هیدرولیکی را تخمین بزنند. به‌عبارت دیگر هر چقدر پراکنش نقاط برآورد شده به خط رگرسیونی ۱:۱ ترسیمی، نزدیک‌تر باشد، این امر بیانگر دقت روش نام برده در انجام تخمین‌های خود بوده است.

با در نظر گرفتن شکل ۲، هر چقدر مقادیر اندازه‌گیری شده و برآورد شده K_s نزدیک به هم باشند، احتمال قرارگیری آن‌ها بر روی خط رگرسیونی ۱:۱ افزایش می‌یابد. همان‌طورکه از شکل برمی‌آید، در روش غیرپارامتریک *k- نزدیک‌ترین همسایه*، تمرکز نقاط برآوردی بر روی خط نام برده، بیش‌تر از سایر روش‌های زمین‌آماری بوده است. به همین ترتیب و در بین روش‌های زمین‌آماری، دو روش کوکریجینگ و کریجینگ ساده توانسته‌اند تخمین‌های بهتری از خود ارایه نمایند. به‌منظور کمی‌سازی دقت تکنیک‌های مختلف ۹ گانه نام برده، از یک سری شاخص‌های آماری استفاده گردید که نتایج آن‌ها در جدول ۲ آمده است. با توجه به داده‌های جدول نام برده، دوباره مشاهده می‌گردد که روش غیرپارامتریک *k- نزدیک‌ترین همسایه* واجد بیش‌ترین دقت و کارایی در انجام تخمین‌های خود بوده است و در واقع عدد ٦٥ درصد برای شاخص کارایی (EF=۰/٦٥) این روش، نشانه توانمند بودن این تکنیک در برآورد مقادیر هدایت هیدرولیکی براساس سایر پارامترهای زودیافت خاک شامل توزیع اندازه ذرات خاک، pH خاک، هدایت الکتریکی عصاره اشباع خاک (EC)، درصد اشباع خاک (SP)، درصد کربن آلی خاک (OC)، مقدار مواد خنثی‌شونده (TNV)، جرم ویژه حقیقی و ظاهری خاک بوده است.

در مرتبه بعدی، روش کوکریجینگ که برای انجام تخمین‌های خود از متغیر جرم ویژه ظاهری خاک به‌عنوان پارامتر کمکی سود برده، توانسته است که دقت به‌نسبت قابل‌قبولی در انجام تخمین‌های خود ارایه دهد.



شکل ۲- هدایت هیدرولیکی اشباع اندازه‌گیری شده و برآورد شده به‌وسیله تکنیک‌های مختلف.



شکل ۳- نقشه پراکنش مقادیر هدایت هیدرولیکی اشباع اندازه‌گیری شده و برآورد شده به‌وسیله تکنیک‌های مختلف.

مقادیر منفی برای آماره CRM، در همه روش‌های استفاده شده در این پژوهش، با در نظر گرفتن شکل آماره، دلالت بر تمایل تمامی روش‌های نام برده در بیش‌برآورد کردن مقادیر تخمینی نسبت به مقادیر اندازه‌گیری شده ضریب هدایت هیدرولیکی دارد.

در مرحله بعد، بر خلاف یکسان بودن پارمترهای ضریب همبستگی (r=۰/۵۹) و کارآیی (EF=۰/٤٦) برای دو روش کریجینگ ساده و کریجینگ جامع، با توجه به اندک بودن خطای تخمین‌ها براساس دو پارامتر ریشه میانگین مربعات خطا (RMSE=٤٧/٦ برای کریجینگ ساده و RMSE=٤٧/٨ برای کریجینگ جامع) و حداکثر خطای تخمینی (ME=٣٤/٠١ برای کریجینگ ساده و ME=٣٤/٨٦ برای کریجینگ جامع)، می‌توان جایگاه سوم را به روش کریجینگ ساده اختصاص داد. سایر روش‌های استفاده شده در این پژوهش نیز با دقت به‌نسبت کم‌تری توانسته‌اند مقادیر ضریب هدایت هیدرولیکی را تخمین بزنند.

یافته‌های این پژوهش در تطابق کامل با نتایج به‌دست آمده از پژوهش‌های نمس و همکاران (۲۰۰۹) بوده و بیانگر توانایی نسبی این روش (K-nn) در مقایسه با سایر روش‌های پارامتریک زمین‌آماری می‌باشد. بر خلاف توانایی این روش در تخمین ویژگی‌های بسیار متغیری مانند ویژگی‌های هیدرولیکی خاک، این پژوهش در داخل کشور و پژوهش‌های چندگانه نمس و همکاران در خارج کشور، جزو اولین و تنها پژوهش‌هایی در این زمینه بوده است. بر این اساس توصیه می‌گردد توانایی این روش در تخمین سایر ویژگی‌های هیدرولیکی با وجود فراهمی یک بانک داده معتبر و بزرگ‌تر نیز مورد آزمایش قرار گیرد.

نتیجه‌گیری

برآورد و تعیین هدایت هیدرولیکی اشباع خاک در تمام مدیریت‌های آبیاری و مدل‌سازی‌های هیدورلوژیکی امری مهم محسوب می‌شود. هدف از این پژوهش، مقایسه ۸ مدل زمین‌آماری با روش غیرپارامتریک *k- نزدیک‌ترین همسایه* برای برآورد ضریب هدایت هیدرولیکی اشباع خاک بوده است. بررسی‌ها نشان داد که تکنیک‌های مختلف ۹ گانه، توانسته‌اند تا حدی، مقادیر ضریب هدایت هیدرولیکی را تخمین بزنند. در روش غیرپارامتریک *k- نزدیک‌ترین همسایه*، تمرکز نقاط برآوردی بر روی خط رگرسیونی ۱:۱ بیش‌تر از سایر روش‌های زمین‌آماری بوده است. در بین روش‌های زمین‌آماری، دو روش کوکریجینگ و کریجینگ ساده توانسته‌اند تخمین‌های بهتری از خود ارایه نمایند. شاخص کارایی (EF=۰/٦٥) روش *k- نزدیک‌ترین همسایه*، نشانه توانمند بودن این تکنیک در برآورد

مقادیر هدایت هیدرولیکی براساس سایر پارامترهای زودیافت خاک شامل توزیع اندازه ذرات خاک، pH خاک، هدایت الکتریکی عصاره اشباع خاک (EC)، درصد اشباع خاک (SP)، درصد کربن آلی خاک (OC)، مقدار مواد خنثی‌شونده (TNV)، جرم ویژه حقیقی و ظاهری خاک می‌باشد.

منابع

1.Biggar, J.W., and Nielsen, D.R. 1976. The spatial variability of the leaching characteristics of the field soil. Water Resour. Res. 12: 78-84.

2.Bouma, J. 1989. Using soil survey data for quantitative land evaluation. Advanced Soil Science, 9: 177-213.

3.Carter, M.R. 1993. Ed. Soil sampling and methods of analysis. Lewis publishers, 823p.

4.Deutsch, C.V. 2002. Geostatistical reservoir modeling. Oxford University Press, 376p.

5.Ghorbani Dashtaki, S., Homaee, M., Mahdian, M.H., and Kouchakzadeh, M. 2009. Site-Dependence Performance of Infiltration Models. Water Resour. Manage. DOI 10.1007/s11269-009-9408-3.

6.Jalali, V.R., and Homaee, M. 2011. Introducing a Nonparametric Model Using k-Nearest Neighbor Technique for Predicting Soil Bulk Density. JWSS-Isfahan University of Technology, 15: 56. 181-191. (In Persian)

7.Jung, W.K., Kitchen, N.R., Sudduth, K.A., and Anderson, S.H. 2006. Spatial Characteristics of Claypan Soil Properties in an Agricultural Field. SSSAJ, 70: 1387-1397.

8.Klute, A. 1986. Ed. Methods of soil analysis, Part 1: Physical and mineralogical methods, Second Edition. Monogr. 9. ASA. and SSSA, Madison,WI. 1188p.

9.Kravchenko, A.N., Bullock, D.G., and Boast, C.W. 2000. JointMultifractal Analysis of Crop Yield and Terrain Slope. Agron. J. 92: 1279-1290.

10.Lall, U., and Sharma, A. 1996. A nearest-neighbor bootstrap for resampling hydrologic time series. Water Resour. Res. 32: 679-693.

11.Lee, M.D., Reynolds, W.D., Elrick, D.E., and Clothier. 1985. A comparison of three field methods for measuring saturated hydraulic conductivity. Can. J. Soil Sci. 65: 563-573.

12.Mallants, D., Mohanty, B.P., Jacques, D., and Feyen, J. 1996. Spatial variability of hydraulic properties in a multi-layered soil profile. Soil Science, 161: 167-181.

13.Miller, E.E., and Miller, R.D. 1956. Physical theory for capillary flow phenomena. J. Appl. Phys. 27: 324-332.

14.Nemes, A., Timlin, D.J., Pachepsky, Ya.A., and Rawls, W.J. 2009. Evaluation of the Rawls *et al.* (1982) Pedotransfer Functions for their Applicability at the U.S. National Scale. Soil Sci. Soc. Am. J. 73: 1638-1645. DOI:10.2136/sssaj 2008.0298.

15.Nemes, A., Roberts, R.T., Rawls, W.J., Pachepsky, Ya.A., and van Genuchten, M.Th. 2008. Software to estimate 33 and 1500 kPa soil water retention using the non-parametric k-Nearest Neighbor technique. Environmental Modelling and Software, 23: 254-255. doi:10.1016/j.envsoft.2007.05.018.
16.Nemes, A., Wösten, J.H.M., Lilly, A., and Oude Voshaar, J.H. 1999. Evaluation of different procedures to interpolate the cumulative particle-size distribution to achieve compatibility within a soil database. Geoderma. 90: 187-202.
17.Page, A.L., Miller, R.H., and Keeney, D.R. 1982. Methods of Soil Analysis, Part II, Physical properties, ASA, SSSA, Madison, WI.
18.Pringle, M.J., Romano, N., Minasny, B., Chirico, G.B., and Lark, R.M. 2007. Spatial evaluation of pedotransfer functions using wavelet analysis. J. Hydrol. 333: 182-198.
19.Rawls, W.J., Brakensiek, D.L., and Saxton, K.E. 1982. Estimation of soil water properties. Transactions of the ASAE. 25: 1316-1320.
20.Reynolds, W.D., and Elrick, D.E. 1987. Laboratory and numerical assessment of the Guelph permeameter method. Soil Sci. 144: 244-282.
21.Schaap, M.G., Leij, F.J., and van Genuchten, M.Th. 2001. Rosetta: a computer program for estimating soil hydraulic parameters with hierarchical pedotransfer functions. J. Hydrol. 251: 163-176.
22.Segal, E., Bradford, S.A., Shouse, P., Lazarovitch, N., and Corwin, D. 2008. Integration of Hard and Soft Data to Characterize Field-Scale Hydraulic Properties for Flow and Transport Studies. Vadose Zone J. 7: 878-889.
23.Sobieraj, J.A., Elsenbeer, H., and Cameron, G. 2004. Scale dependency in spatial patterns of saturated hydraulic conductivity. Catena, 55: 49-77.
24.Twarakavi, N.K.C., Šimůnek, J., and Schaap, M.G. 2009. Development of Pedotransfer Functions for Estimation of Soil Hydraulic Parameters using Support Vector Machines. Soil Sci. Soc. Am. J. 73: 1443-1452.
25.Verzani, J. 2004. Using R for Introductory Statistics. Chapman and Hall/CRC Press, 432p.
26.Walkley, A., and Black, I.A. 1934. An examination of the Degtjareff for determining soil organic matter and a proposed modification of the chromic acid titration method. Soil Science, 37: 29-38.
27.Webster, R., and Oliver, M.A. 2007. Geostatistics for Environmental Scientists. John Wiley & Sons Ltd. 332p.
28.Zeleke T.B., and Si, B.C. 2005. Scaling Relationships between Saturated Hydraulic Conductivity and Soil Physical Properties. SSSA J. 69: 1691-1702.

Gorgan University of Agricultural Sciences and Natural Resources

J. of Water and Soil Conservation, Vol. 20(5), 2013
http://jwsc.gau.ac.ir

Comparing geostatistics techniques and nonparametric *k-nearest neighbor* technique for predicting soil saturated hydraulic conductivity

***V.R. Jalali[1], A. Khashei Siouki[2] and M. Homaee[3]**

[1]Assistant Prof., Dept. of Soil Science, Shahid Bahonar University of Kerman,
[2]Assistant Prof., Dept. of Water Engineering, University of Birjand,
[3]Professor, Dept. of Soil Science, Tarbiat Modares University
Received: 07/21/2012; Accepted: 11/05/2012

Abstract

Knowledge of the soil saturated hydraulic conductivity (K_s) is essential for irrigation management purposes and hydrological modeling, but it cannot often be measured because of practical and/or cost-related reasons. In this research, common geostatistical approaches with one type of the nonparametric lazy learning algorithms, a k-nearest neighbor (k-NN) algorithm, was compared and tested to estimate saturated hydraulic conductivity (Ks) from other easily available soil properties. In this research 151 soil samples were collected from arable land around Bojnourd and saturated hydraulic conductivity (Ks) was estimated from other soil properties including soil textural fractions, EC, pH, SP, OC, TNV, ρs and ρb. The nonparametric k-NN technique performed mostly equally well, in terms of Pearson correlation coefficient (r), modeling efficiency (EF), root-mean-squared errors (RMSE), maximum error (ME) and coefficient of residual mass (CRM) statistics (r=0.76, EF=0.655, RMSE=42.87, ME=26.89 and CRM= -0.11) and after that, Co-Kriging and simple kriging methods, performed better than others. It can be concluded that the k-NN technique is a competitive alternative to other techniques such as pedotransfer functions (PTFs) to estimate saturated hydraulic conductivity especially when for new data set redriving these functions is essential.

Keywords: Geostatistics techniques, *K-nearest neighbor* technique, Soil saturated hydraulic conductivity

* Corresponding Author; Email: v.jalali@uk.ac.ir