

ارزیابی روش غیرپارامتریک k - نزدیک‌ترین همسایه و سیستم‌های شبکه عصبی مصنوعی برای برآورد هدایت هیدرولیکی اشباع خاک

*عباس خاشعی‌سیوکی^۱، وحیدرضا جلالی‌موخر^۲، علی محمد نوفرستی^۳ و یوسف رضانی^۱

استادیار گروه مهندسی آب، دانشگاه بیرجند، استادیار گروه علوم خاک، دانشگاه شهید باهنر کرمان، مربی گروه مهندسی آب، دانشگاه بیرجند
تاریخ دریافت: ۹۲/۱۰/۹؛ تاریخ پذیرش: ۹۳/۷/۶

چکیده

سابقه و هدف: هدایت هیدرولیکی اشباع خاک از مهم‌ترین ویژگی‌های فیزیکی خاک است ولی در بیش‌تر موارد به علت محدودیت‌های عملی و یا هزینه‌ای، اندازه‌گیری آن با دشواری همراه است. در این پژوهش مدل‌های مختلف شبکه‌های عصبی مصنوعی 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. **مواد و روش‌ها:** در این پژوهش ۱۵۱ نمونه از خاک‌های زراعی اطراف بجنورد، انتخاب 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) و میزان مواد خنثی‌شونده آن (TNV) جهت برآورد هدایت هیدرولیکی اشباع به کار گرفته شد. جهت ارزیابی سیستم‌های شبکه عصبی مصنوعی همه داده‌ها به ۳ قسمت، ۵۰ درصد برای آموزش، ۲۵ درصد برای تست و سایر داده‌ها برای اعتبارسنجی تقسیم شدند. طراحی ساختار مدل‌های MLP (پرسپترون چندلایه) توسط توابع سیگموئید و لایه مخفی انجام شد. شبکه عصبی مصنوعی برای همه مدل‌ها، با الگوریتم آموزشی لوبنبرگ-مارکوآردت به صورت یک لایه پنهان تابع آستانه $\log\text{sig}$ برای لایه پنهان و tansig برای لایه خروجی انتخاب گردید.

یافته‌ها: استفاده از پارامترهای آماری نشان داد که از لحاظ دقت برآورد، روش شبکه عصبی مصنوعی در مقایسه با روش غیرپارامتریک k - نزدیک‌ترین همسایه در شرایط ارائه تمامی پارامترها (با داشتن آماره‌های $EF=0/946$ ، $RMSE=8/798$ ، $ME=28/446$ و $CRM=-0/134$) نسبت به سایر روش‌ها و مدل‌های ورودی از دقت قابل قبولی برخوردار می‌باشد.

نتیجه‌گیری: بررسی‌ها نشان داد که تکنیک‌های مختلف 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 مورد بررسی بوده است. بهترین نتیجه مربوط به روش شبکه عصبی مصنوعی با تمامی اطلاعات بانک داده بود شاخص کارایی (۵۷ تا ۷۱ درصد EF) روش k - نزدیک‌ترین همسایه، نشانه توانمند بودن این تکنیک در برآورد مقادیر هدایت هیدرولیکی براساس سایر پارامترهای زودیافت خاک شامل توزیع اندازه ذرات

* مسئول مکاتبه: abbaskhashei@birjand.ac.ir

خاک، هدایت الکتریکی عصاره اشباع خاک (EC)، درصد اشباع خاک (SP)، درصد کربن آلی خاک (OC)، مقدار مواد خنثی شونده (TNV)، جرم ویژه حقیقی و ظاهری خاک می‌باشد. روش شبکه عصبی مصنوعی می‌تواند به‌عنوان روشی جایگزین برای اشتقاق توابع انتقالی خاک، به‌ویژه هنگامی که فراهمی داده‌های جدید، نیاز به اشتقاق مجدد این توابع را الزام‌آور می‌کند، به‌کار رود.

واژه‌های کلیدی: روش‌های شبکه عصبی مصنوعی، تکنیک k -نزدیک‌ترین همسایه، هدایت هیدرولیکی اشباع

مقدمه

هدایت هیدرولیکی اشباع از مهم‌ترین مشخصات هیدرولیکی خاک می‌باشد و به‌صورت گسترده در اندازه‌گیری و ارزیابی کیفیت فیزیکی خاک و به‌طور کلی در مطالعه جریان آب در خاک مورد استفاده قرار می‌گیرد. این پارامتر در طرح‌های آب و خاک، طراحی شبکه‌های آبیاری و زهکشی، پژوهش‌های جریان آب در سفره‌های زیرزمینی، جریان آب در سدهای خاکی و همچنین در آبخویی به‌منظور نمک‌زدایی و دفع مواد آلوده‌کننده خاک کاربرد دارد (۳). یکی از مشکلات مهمی که کاربرد یافته‌های مربوط به ویژگی‌های هیدرولیکی خاک را در مورد پیش‌بینی و ارزیابی فرآیندهای موجود در محیط متخلخل خاک تحت‌تأثیر قرار می‌دهد، تغییرپذیری مکانی^۱ توابع هیدرولیکی خاک است. راه‌حل‌هایی متفاوت برای حل این مشکل توسط پژوهشگران مختلف ارائه شده است که از آن جمله می‌توان به همبستگی تجربی^۲ بین متغیرهای خاکی مانند توابع انتقالی (۱)، استفاده از نظریه مقیاس محیط‌های مشابه^۳ (۱۰)، استفاده از تئوری متغیرهای ناحیه‌ای و روش‌های زمین‌آماري^۴ (۷) 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.

آگاهی از هدایت آبی اشباع خاک برای درک و مدل‌سازی بسیاری از فرآیندهای فیزیکی خاک ضروری است (۲۲). 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 جهت حل این مشکل، روش‌های غیرمستقیم را مورد توجه قرار داده‌اند تا به‌وسیله آن بتوانند در نقاط فاقد داده، تخمینی بهینه از متغیر موردنظر داشته باشند. اشتقاق توابع انتقالی خاک و استفاده از روش‌های هوش مصنوعی از جمله این روش‌هاست (۱). تکنیک‌های رگرسیونی و اخیراً، شبکه‌های عصبی مصنوعی دو روش معمول در توسعه توابع انتقالی خاک می‌باشند. شبکه‌های عصبی مصنوعی همان‌گونه که از نام آن پیداست، تکنیکی برای مغز انسان در فرایند حل مسأله است. انسان‌ها از دانش به‌دست آمده از تجربیات برای حل مسائل یا شرایط جدید استفاده می‌کنند، ساختار شبکه‌های عصبی می‌تواند برای محاسبات قدرتمند از روابط پیچیده غیرخطی استفاده شود. اسشاپ و لیچ (۱۹۹۸)، ۱۹ مدل شبکه عصبی مصنوعی برای پیش‌بینی هدایت هیدرولیکی اشباع با استفاده از پارامترهای درصد شن، سیلت و رس، جرم ویژه ظاهری، تخلخل کل، درصد سنگ‌ریزه، افق خاک و نقاط منحنی رطوبتی ایجاد کردند (۱۷). بررسی نصرتی‌کاریزک و همکاران (۲۰۱۲) نشان داد که بر پایه نتایج تجزیه حساسیت، درصد شن و رس، درصد

- 1- Spatial Variability
- 2- Experiential Correlation
- 3- Similar media Scaling Theory
- 4- Geostatistical Techniques

انتقالی از دقت کافی جهت استفاده آن‌ها در مقیاس ایالات متحده آمریکا برخوردار نبوده و با ارزیابی روش غیرپارامتریک k - نزدیک‌ترین همسایه، بیان نمودند که روش مذکور از توانایی بالاتری جهت تخمین توابع هیدرولیکی در مقیاس کل ایالات متحده آمریکا برخوردار است (۱۱).

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 مذکور، نمس و همکاران (۱۹۹۹) با استفاده از تکنیک k - نزدیک‌ترین همسایه، به تخمین ویژگی‌های هیدرولیکی خاک پرداختند و بیان نمودند که علی‌رغم دقت یکسان تکنیک مذکور و روش شبکه‌های عصبی مصنوعی در اشتقاق و تخمین توابع هیدرولیکی، قابلیت روش k - نزدیک‌ترین همسایه جهت وارد نمودن داده‌های محلی، ارجحیتی نسبی برای این روش ایجاد می‌نماید البته آن‌ها از تمامی اطلاعات ایالات متحده استفاده کردند و مدل‌ها را با داده‌های کم مورد تحلیل قرار ندادند (۱۲). رسول‌زاده و همکاران (۲۰۱۲) پژوهشی جهت، ارزیابی مدل‌های توابع انتقالی رگرسیونی مختلف، مدل تابع انتقالی رزتا با ماهیت شبکه عصبی مصنوعی و مدل فرکتال برای برآورد هدایت هیدرولیکی اشباع انجام دادند (۱۴).

سیگال و همکاران (۲۰۰۸) نیز به مطالعه حرکت و انتقال آب و املاح در مقیاس ناهمگون مزرعه‌ای پرداختند. ایشان در پژوهش خود از تکنیک k - نزدیک‌ترین همسایه به‌عنوان الگوریتم میان‌یابی نام برده و به‌طور موفقیت‌آمیزی توانستند با استفاده از این رویکرد، درک عمیقی از تغییرپذیری ویژگی‌های هیدرولیکی خاک در مقیاس مزرعه‌ای به‌دست آورند (۱۶).

رطوبت در مکش ۳ بار، تخلخل کل و میانگین هندسی اندازه ذرات حساس‌ترین پارامترها در برآورد هدایت هیدرولیکی اشباع بودند (۱۳). اسخاپ و همکاران (۲۰۰۱) با استفاده از نرم‌افزار رزتا^۱ (توابعی انتقالی براساس شبکه‌های عصبی مصنوعی)، اقدام به تخمین و برآورد ویژگی‌های هیدرولیکی خاک نمودند (۱۸).

وجه تشابه بین اکثر توابع انتقالی موجود، در اشتقاق آن‌ها بر مبنای رویکرد پارامتریک می‌باشد. بدین 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. استفاده از تکنیک‌های غیرپارامتریک می‌تواند به‌عنوان رویکردی جایگزین برای چنین 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 رزتا را براساس رویکرد پارامتریک و با استفاده از شبکه‌های عصبی مصنوعی ابداع نموده بودند (۱۸). در پژوهشی نوین در سال 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 (۱۹)، ولی در پژوهش فوق روش‌های ناپارامتریک نزدیک‌ترین همسایه K -nn^۲ مورد تحلیل قرار نگرفت. 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

1- Rosetta
2- K-Nearest Neighbor

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 (۸).

همچنین سایر ویژگی‌های خاک شامل، هدایت الکتریکی عصاره اشباع خاک (EC_e)، درصد مواد آلی خاک (OM) والکلی و بلک، (۱۹۳۴) (۱۹)، رطوبت اشباع خاک (θ_s) و میزان مواد خنثی‌شونده آن نیز (TNV)^۱ (۸) تعیین گردید. در مرحله بعد، با استفاده از رابطه لال و شارما (۱۹۹۶)، تخمینی اولیه از تعداد k نمونه جهت وارد نمودن در محاسبات به‌عمل آمد و سپس با استفاده از تکنیک ارزیابی تقاطعی، تعداد دقیق k نمونه خاک محاسبه گردید (۱۱).

تکنیک اعتبارسنجی متقابل^۲ 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^۳ 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 $n+1$ تعداد از داده‌های ورودی محاسبه می‌کند. این چرخش تا زمانی ادامه می‌یابد تا سرانجام، ترکیب مورد نظر که دارای کم‌ترین خطا بین مقادیر تخمینی و مشاهده‌ای است، به‌دست آید.

پس از تعیین تعداد k - نزدیک‌ترین همسایه جهت ورود به محاسبات، برنامه موردنظر جهت ورود ویژگی‌های هر خاک شامل مختصات جغرافیایی هر نقطه در سیستم متریک، درصد ذرات شن، سیلت و رس، درصد مواد خنثی‌شونده (TNV)، هدایت

ذولفقاری و همکاران (۲۰۱۲) به مقایسه روش‌های k - نزدیک‌ترین همسایگی و شبکه عصبی مصنوعی در برآورد ظرفیت تبادل کاتیونی خاک پرداختند همچنین نتایج بررسی کارایی مدل‌ها نشان داد که هر دو روش از کارایی بالایی ($EF=0/88$) در برآورد ظرفیت تبادل کاتیونی خاک برخوردار هستند (۲۳).

بر این اساس هدف اصلی این پژوهش، مقایسه کارایی نسبی روش غیرپارامتریک $K-nn$ با روش‌های سیستم‌های خبره شبکه عصبی مصنوعی در تخمین ضریب هدایت هیدرولیکی اشباع خاک براساس سایر متغیرهای در دسترس بوده است و کارایی این دو مدل را در برآورد هدایت هیدرولیکی اشباع با بانک اطلاعاتی محدود بررسی می‌نماید.

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. در آذر سال ۲۰۰۹ از بخش‌های مذکور نیز از عمق ۰-۲۵ سانتی‌متری نمونه‌برداری خاک انجام شد و تعداد ۱۵۱ نمونه خاک با شرایط متنوع انتخاب گردید.

در این پژوهش از دستگاه نفوذسنج گلف مدل Eijkelkamp هلند برای تعیین ضریب آبگذری اشباع خاک استفاده گردید. روش نفوذسنج گلف (۱۵) 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 (۸) تعیین شد.

1- Total Neutralizing Value
2- Cross Validation
3- Rotation estimation

که در آن، X : مقدار اولیه X_{min} حداقل و X_{max} حداکثر و X_{norm} : مقدار نرمال شده و $0/8$ و $0/1$ عوامل سنجش هستند. برای عوامل سنجش مقادیر مختلفی ممکن است تعیین شود. با این حال هیچ قانون پیشنهادی و روش استاندارد وجود ندارد که بتواند در شرایط خاص مورد استفاده قرار بگیرد. در این مطالعه عوامل سنجش به ترتیب $0/8$ و $0/1$ انتخاب شدند (۵). همان طور که در جدول‌های ۱ و ۲ ارائه شده است خلاصه آماری داده‌های مورد استفاده و نوع مدل ورودی هر کدام از سیستم‌های شبکه عصبی و مدل ناپارامتریک نشان داده شده است. جدول ۱، توصیفی آماری از ویژگی‌های انتخابی خاک‌های موجود در بانک داده را نشان می‌دهد.

در نهایت اقدام به ارزیابی و اعتبارسنجی عملکرد مدل‌ها با استفاده از یک سری شاخص‌های آماری شد. یکی از شاخص‌های آماری که برای ارزیابی مدل‌ها از آن استفاده می‌شود، ضریب همبستگی پیرسون می‌باشد

قضاوت از روی این ضریب ساده است و ممکن است به نظر برسد که ضریب همبستگی می‌تواند معیار مناسبی در ارزیابی مدل باشد. با این حال 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 به گونه‌ای باشد که از یک روند یکنواخت پیروی نماید. بنابراین اگرچه ضریب همبستگی به خوبی نشان‌دهنده میزان هماهنگی روند تغییرات مقادیر مشاهده شده نسبت به مقادیر پیش‌بینی شده می‌باشد اما گویای تطابق آن‌ها نیست (۴).

الکتريکی 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.

جهت ارزیابی سیستم‌های شبکه عصبی مصنوعی^۱ همه داده‌ها به ۳ قسمت، ۵۰ درصد برای آموزش^۲، ۲۵ درصد برای تست و سایر داده‌ها برای اعتبارسنجی^۳ تقسیم شدند. طراحی ساختار مدل‌های^۴ MLP (پرسپترون چندلایه) توسط توابع سیگموئید و لایه مخفی انجام شد. شبکه عصبی مصنوعی برای همه مدل‌ها، با الگوریتم آموزشی لونیبرگ-مارکوآردت به‌صورت یک لایه پنهان تابع آستانه logsig برای لایه پنهان و tansig برای لایه خروجی انتخاب گردید. در بررسی نصرتی‌کاريزک و همکاران (۲۰۱۲) تعداد بهینه سلول‌های عصبی مخفی توسط روش آزمون و خطا انتخاب شد و برای جلوگیری از بیش برآورد^۵، یک معیار توقف به موقع به‌منظور بهبود سرعت و کارایی آموزش شبکه تعیین شد. برای اجرای مدل شبکه عصبی مصنوعی از بخش ANN ، toolbox در نرم‌افزار MATLAB استفاده شد (۵). به‌منظور به‌دست آوردن پایداری مدل، مجموعه داده‌ها برای بار اول در محدوده ۰ تا ۱ نرمال شدند و سپس به مقادیر اولیه برگشتند (۲).

$$X_{norm} = \left[\frac{X - X_{min}}{X_{max} - X_{min}} \right] \times 0.8 + 0.1 \quad (1)$$

- 1- Artificial Neural Network
- 2- Training
- 3- Validation
- 4- Multilayered Perceptrons
- 5- Over-fitting

جدول ۱- خلاصه‌ای از آماره‌های توصیفی پارامترهای خاکی به کار رفته جهت تخمین k - نزدیک‌ترین همسایه.
 Table 1. A summary of descriptive statistics soil parameters used to estimate the k-nearest neighbor.

ضرب تغییرات C.V	انحراف معیار Standard Division	میانگین Average	بیشینه Max	کمینه Min	دامنه Range	واحد Unit	ویژگی properties
5.62	0.086	1.53	1.75	1.26	0.49	g.cm ⁻³	جرم مخصوص ظاهری Bulk Density
3.43	0.088	2.56	2.78	2.19	0.59	g.cm ⁻³	جرم مخصوص حقیقی
39.77	0.08283	0.2107	0.75	0.04	0.71	g.g ⁻¹	شن Sand
11.1	0.05408	0.4870	0.62	0.13	0.49	g.g ⁻¹	سیلت Silt
15	0.04482	0.3023	0.44	0.12	0.32	g.g ⁻¹	رس Clay
24.9	0.22	0.88	1.91	0.21	1.7	%	کربن آلی OC
24.38	5.36	21.99	47.75	5.75	42	%	مواد خنثی شونده TNV
28	0.48	1.74	3.81	0.27	3.54	dS.m ⁻¹	شوری EC
3.1	0.0126	0.407	0.46	0.38	0.09	m ³ .m ⁻³	رطوبت اشباع SP(θ_s)
1.16	19.85	17.12	194.62	1.39	193.22	cm.d ⁻¹	هدایت هیدرولیکی اشباع خاک Ks

جدول ۷- پارامترهای مورد استفاده در هر یک از مدل‌های سیستم‌های خبره و ناپارامتریک.

نوع مدل ورودی Type of inputs	طول Longitude	عرض Latitude	درصد شن sand Percent	درصد سیلت Silt percent	درصد رس Clay percent	کربن آلی Organic Carbon OC	مواد خنثی‌شونده (TNV) Total Neutralizing Value	شوری Electric Conductivity	درصد آهک Percentage of lime	رطوبت اشباع Saturation (θ_s) moisture	جرم ویژه ظاهری Bulk Density	جرم ویژه حقیقی Particle density
a	*	*	*	*	*	*						
b	*	*	*	*	*	*	*					
c	*	*	*	*	*	*	*	*	*	*		
d	*	*	*	*	*	*	*	*	*	*	*	*

موجود در بانک مرجع ۱۵۱ ($n > 100$) عدد بود، بنابراین:

$$k = n^{1/2} \text{ for } n > 100$$

$$k = 151^{1/2} = 12.3$$

عدد به دست 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)^۵ نشان می‌دهد. همان‌طور که مشاهده می‌شود، مقدار خطا در کم‌ترین تعداد همسایگی یعنی $K=1$ حداکثر بوده و با افزایش این تعداد از میزان خطا کاسته شده است. این روند تا $K=10$ ادامه یافته ولی پس از آن دوباره میزان خطا افزایش یافته و یا به عبارت دیگر دقت تخمین‌ها کاهش یافته است. پس در واقع در دامنه مورد نظر (۱ تا ۱۵) تعداد $K=10$ بهینه‌ترین تعداد همسایگی جهت انجام تخمین‌ها بوده است.

شاخص‌های کمی دیگری که می‌توان در برآورد دقت مدل از آن‌ها استفاده نمود، عبارتند از آماره‌های خطای ماکزیمم (ME)^۱، ریشه میانگین مربعات خطا (RMSE)^۲، کارایی مدل (EF)^۳ و ضریب جرم باقی‌مانده (CRM)^۴. بیان ریاضی آماره‌های مذکور به صورت زیر است:

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

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

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

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

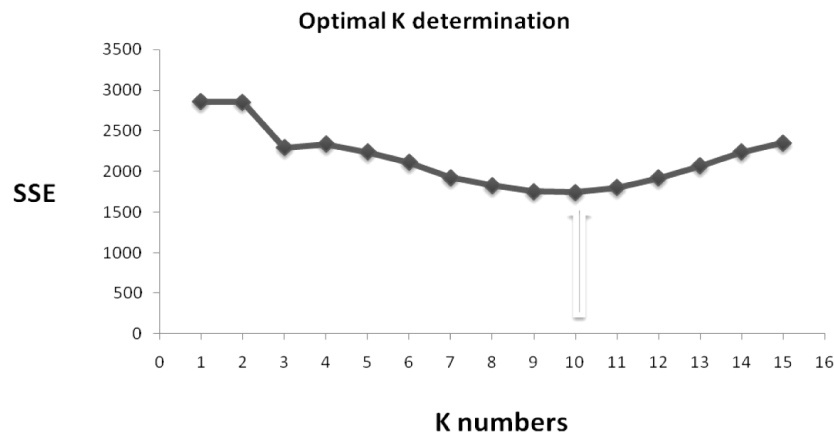
اگر تمامی داده‌های برآورد شده و اندازه‌گیری شده یکسان باشند، نتایج این آماره‌ها به صورت $ME=0$ ، $RMSE=0$ ، $EF=1$ و $CRM=0$ خواهد بود.

نتایج و بحث

همان‌گونه که اشاره شد، رابطه لال و شارمال (۱۹۹۶) تخمینی اولیه از مقدار k تعداد بهینه از نزدیک‌ترین همسایه‌ها از بانک مرجع به داده هدف ارائه می‌دهد (۶). با توجه به این‌که تعداد داده‌های

5- Sum of Square Error

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



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

Figure 1. Determine of the optimal k based on sum of squared error (SSE).

نشان داد که بهترین مدل از نظر دقت و سرعت تخمین هدایت هیدرولیکی اشباع مدلی بود که از پارامترهای ورودی لگاریتم میانگین هندسی قطر ذرات، تخلخل کل و درصد شن و رس استفاده شد. نتایج پژوهش فوق با نتایج مطالعه نصرتی و همکاران (۲۰۱۲) تا حدودی مشابه است البته ایشان در پژوهش فوق متغیر مکانی را در نظر نگرفته‌اند و تنها به متغیرهای فیزیکی زودیاقت خاک بسنده کرده‌اند و به همین دلیل نمی‌توان به صورت دقیق نتایج را مورد مقایسه قرار داد ولی هر دو مدل بر تأثیر داده‌های فیزیکی خاک بر پیش‌بینی هدایت هیدرولیکی خاک اتفاق نظر دارند (۱۳).

بررسی‌ها نشان داد که مدل c از بین مدل‌های غیرپارامتریک بالاترین دقت را دارا است با توجه به این که میزان ضریب همبستگی پیرسون کم‌تر از مدل d در روش غیرپارامتریک است ولی این ورودی توانسته است خطای کم‌تری در برآورد داشته باشد و دارای CRM ، ME و $RMSE$ کم‌تری نسبت به مدل d باشد. میزان کاهش خطا در مدل c نسبت به d ۳۵ درصد کاهش یافته است. مدل c در مقایسه با مدل

در مرحله بعد الگوریتم موردنظر در محیط برنامه R نوشته شد (۶) 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. جدول ۳ میزان هر کدام از این آماره‌ها را نشان می‌دهد.

با توجه با داده‌های جدول مشاهده می‌شود که روش شبکه عصبی مصنوعی با تمام داده‌ها یعنی مدل d توانسته است با $r=0/97$ و $RMSE=8/89$ بهترین مدل در بین مدل‌ها مورد آزمون باشد. همان‌طور که از جدول ۳ مشخص است مقدار $EF=0/946$ بوده و به یک نزدیک است. این نتیجه با نتایج پژوهش‌های اسشاپ و همکاران (۲۰۰۱) اس شاپ و لیچ (۱۹۹۸) نیز مطابقت دارد (۱۶، ۱۷). بهترین مدل براساس رتبه‌بندی تعیین شده بدین صورت که که مدلی که بتواند از تمامی آماره‌های آزمون برخوردار باشد مثلاً R^2 بالا و خطای کم‌تری داشته باشد، بهترین مدل می‌باشد. نتایج نصرتی‌کاریزک و همکاران (۲۰۱۲)

آهک) نمی‌تواند از دقت قابل‌قبولی برخوردار باشد، علت آن می‌تواند مربوط به همبستگی پایین بین داده‌های رطوبت اشباع و میزان شوری با هدایت هیدرولیکی اشباع در بانک اطلاعاتی مورد استفاده باشد (به ترتیب ۰/۰۰۱، ۰/۰۲۸، ۰/۰۳).

مشابه روش شبکه عصبی مصنوعی نیز از دقت مناسب‌تری برخوردار است به طوری که باعث بهبودی ۲۰ درصدی در ضریب همبستگی پیرسون بوده است. بررسی‌ها نشان داد که ورودی c علی‌رغم بالا بودن تعداد نسبی ورودی (طول و عرض جغرافیایی، درصد شن، سیلت و رس و درصد آهک، درجه اشباع و

جدول ۳- آماره‌های محاسباتی جهت تعیین میزان قابلیت تکنیک‌های مختلف در برآورد هدایت هیدرولیکی خاک.

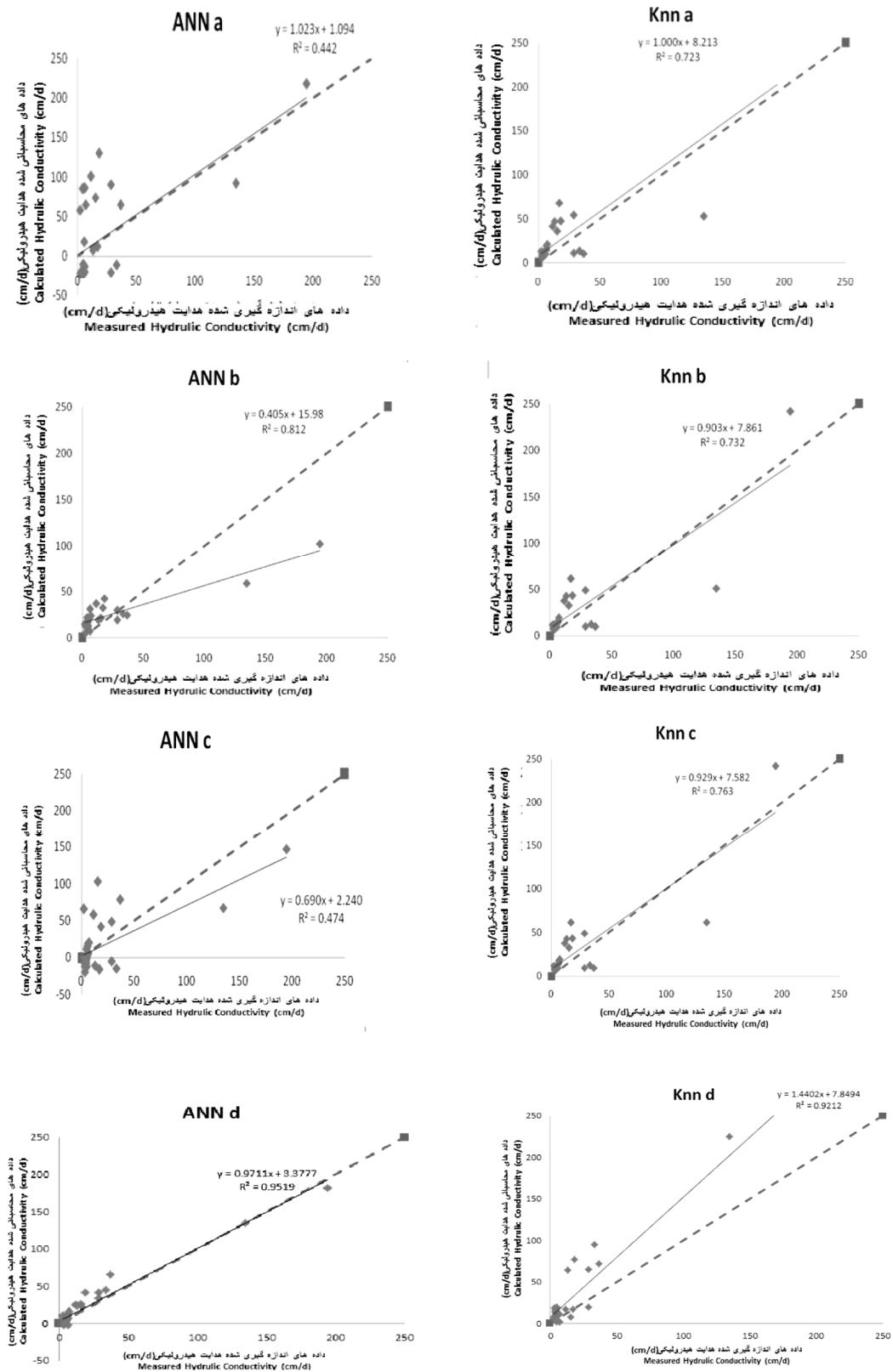
Table 3. Calculating statistics to determine the ability of different techniques for estimating soil hydraulic conductivity.

رتبه مدل Rank model	تعداد اپوچ Epoch Number	RMSE	r	CRM	ME	EF	مدل ورودی Input Model	نوع سیستم system type
8	15	43.44	0.665	-0.076	111.79	-0.321	a	شبکه عصبی مصنوعی Artificial Neural Network
3	18	24.193	0.901	-0.216	92.37	0.59	b	
7	19	30.704	0.688	0.231	87.47	-5.353	c	
1	12	8.79	0.976	-0.134	28.447	0.946	d	
5	-	68.305	0.850	-0.44	81.74	0.570	a	روش غیر پارامتریک Non-parametric method
4	-	60.180	0.856	-0.33	83.71	0.667	b	
2	-	57.09	0.86	-0.341	73.01	0.7	c	
6	-	88.99	0.96	-0.491	102.76	0.71	d	

خوبی مقادیر هدایت هیدرولیکی اشباع خاک را برآورد نماید ($R^2=0/72$).

به‌طور کلی می‌توان گفت که با در نظر گرفتن داده‌های جدول ۳، تکنیک‌های مختلف مورد بررسی، توانسته‌اند تا حدی، مقادیر ضریب هدایت هیدرولیکی را تخمین بزنند. بارت دیگر هر چقدر پراکنش نقاط برآورد شده به خط رگرسیونی ۱:۱ ترسیمی، نزدیک‌تر باشد، این امر بیانگر دقت روش مذکور در انجام تخمین‌های خود بوده است.

در شکل ۲ مقادیر پیش‌بینی با مقادیر محاسباتی هدایت هیدرولیکی اشباع در یک نمودار برای تمامی مدل‌ها ارائه شده است. با توجه به 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 غیر پارامتریک k - نزدیک‌ترین همسایه، روند داده‌های اندازه‌گیری شده و برآورد شده برای ورودی a پراکنش کم‌تری از خط ۱:۱ نسبت به مدل شبکه عصبی داشته است یا به‌عبارت دیگر مدل مذکور توانسته است با دقت نسبتاً



شکل ۲- مقایسه خروجی مدل‌های ANN و knn با داده‌های مشاهداتی و خط رگرسیونی ۱:۱.

Figure 2. Compares the output knn and ANN models with observed data and linear regression 1:1.

کاری مشکل است اما می‌توان در صورت در اختیار داشتن تعداد معدودی از داده‌ها مدل ناپارامتریک k - نزدیک‌ترین همسایه را به‌عنوان یکی از روش‌های مناسب برآورد پارامترها مشکل از روی داده‌های آسان‌تر به‌کار برد.

نتایج خروجی مدل‌های شبکه عصبی نشان می‌دهد که بیش‌تر خروجی‌های بالاتر از خط ۱:۱ است و این امر نشان از کم برآورد بودن مدل‌های شبکه عصبی دارد. این در حالی است که مدل غیرپارامتریک نزدیک‌ترین همسایه نیز در بیش‌تر مواقع میزان تخمینی پارامتر را کم‌تر از مقدار واقعی تخمین زده است.

در تخمین ویژگی‌های بسیار متغیری مانند ویژگی‌های هیدرولیکی خاک، این پژوهش در داخل و خارج کشور، جزو اولین و تنها پژوهش‌هایی در این زمینه بوده است که در آن دو مدل با بانک اطلاعاتی کم مورد بررسی و مقایسه قرار می‌گیرند، علی‌رغم این‌که در ارتباط با خود مدل k - نزدیک‌ترین همسایه و روش شبکه عصبی به‌صورت مجزا پژوهش‌هایی انجام شده 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 هیدرولوژیکی امری مهم محسوب می‌شود. هدف از این پژوهش، مقایسه مدل ورودی شبکه عصبی مصنوعی با همین ورودی‌ها با روش غیرپارامتریک k - نزدیک‌ترین همسایه جهت برآورد ضریب هدایت هیدرولیکی اشباع خاک می‌باشد. بررسی‌ها نشان داد که تکنیک‌های مختلف توانسته‌اند تا حدی، مقادیر

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. همان‌طور که از شکل برمی‌آید، در روش غیرپارامتریک k - نزدیک‌ترین همسایه، تمرکز نقاط برآوردی بر روی خط مذکور، بیش‌تر برای ورودی‌ها با تعداد کم‌تر است در شکل نیز مشخص است که خط برازش داده شده به خط ۱:۱ نزدیک‌تر است نکته قابل‌توجه این است که مدل ناپارامتریک نسبت به مدل شبکه عصبی برای ورودی‌های زیاد از دقت کم‌تری برخوردار است.

با توجه به داده‌های جدول ۳، دوباره مشاهده می‌گردد که روش غیرپارامتریک k - نزدیک‌ترین همسایه واجد بیش‌ترین دقت و کارایی در انجام تخمین‌های خود بوده است و در واقع عدد ۵۷ تا ۷۱ درصد برای شاخص کارایی (EF) این روش، نشانه توانمند بودن این تکنیک در برآورد مقادیر هدایت هیدرولیکی براساس سایر پارامترهای زودیافت خاک شامل توزیع اندازه ذرات خاک، هدایت الکتریکی عصاره اشباع خاک (EC)، رطوبت اشباع خاک، درصد کربن آلی خاک (OC)، مقدار مواد خنثی‌شونده (TNV)، جرم ویژه حقیقی و ظاهری خاک برای همین پارامترها برای شبکه عصبی مصنوعی است.

در مرتبه‌بندی (جدول ۲)، روش شبکه عصبی با کم‌ترین داده ورودی دارای آخرین رتبه بود که علت آن بالا بودن $ME=111/75$ و پایین بودن شاخص کارایی $EF=-0/321$ می‌باشد. یافته‌های این پژوهش در تطابق کامل با نتایج حاصل از پژوهش‌ها نمس و همکاران (۱۹۹۹) بوده و بیانگر توانایی نسبی روش (K-nn) در مقایسه با روش هوشمند شبکه عصبی مصنوعی به‌ازای ورودی‌ها کم می‌باشد (۱۲). علی‌رغم توانایی روش شبکه عصبی به‌ازای ورودی‌های زیاد این امر به‌علت مشکل بودن اندازه‌گیری تمامی داده‌ها

(TNV)، جرم مخصوص حقیقی و ظاهری خاک می‌باشد.

سپاسگزاری

داده‌های استفاده شده در این پژوهش حاصل طرح تحقیقاتی شماره ۱۱/۲۵۸۷۲ با عنوان "ایجاد بانک داده خاک تحت وب؛ ابزاری مناسب جهت پایش ویژگی‌های خاک در مقیاس ملی" و با حمایت مالی معاونت علمی و فناوری ریاست جمهوری (ستاد توسعه فناوری آب، خشکسالی، فرسایش و محیط زیست) انجام پذیرفته که بدین‌وسیله قدردانی می‌گردد.

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 مورد بررسی بوده است. بهترین نتیجه مربوط به روش شبکه عصبی مصنوعی با تمامی اطلاعات بانک داده بود شاخص کارایی (۵۷ تا ۷۱ درصد) EF روش k -نزدیک‌ترین همسایه، نشانه توانمند بودن این تکنیک در برآورد مقادیر هدایت هیدرولیکی براساس سایر پارامترهای زودیافت خاک شامل توزیع اندازه ذرات خاک، هدایت الکتریکی عصاره اشباع خاک (EC)، درصد اشباع خاک (SP)، درصد کربن آلی خاک (OC)، مقدار مواد خثی‌شونده

منابع

1. Bouma, J. 1989. Using soil survey data for quantitative land evaluation. *Advanced Soil Science*. 9: 177-213.
2. Doğan, E. 2008. Reference Evapotranspiration Estimation Using Adaptive Neuro-Fuzzy Inference System. *J. Irrig. Drain. Eng. ASCE*. 58: 617-628.
3. Ghorbani Dashtaki, Sh., Homaei, M., and Mahdian, M.H. 2009. Estimating soil water infiltration parameters using Artificial Neural Networks. *J. Water Soil*. 23: 1. 185-198. (In Persian)
4. Jalali, V.R., and Homaei, 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)
5. Jung, W.K., Kitchen, N.R., Sudduth, K.A., and Anderson, S.H. 2006. Spatial Characteristics of Claypan Soil Properties in an Agricultural Field. *Soil Sci. Soc. Amer. J.* 70: 1387-1397.
6. Khashei-Siuki, A., Kouchakzadeh, M., and Ghahraman, B. 2011. Predicting dryland wheat yield from meteorological data using expert system in Khorasan province, I.R. Iran, *JAST*. 13: 627-640.
7. Klute, A. 1986. *Methods of soil analysis, Part 1: Physical and mineralogical methods*, Second Edition. Monogr. 9. ASA. and SSSA, Madison, WI. 1188p.
8. Lall, U., and Sharma, A. 1996. A nearest-neighbor bootstrap for resampling hydrologic time series. *Water Resour. Res.* 32: 679-693.
9. Miller, E.E., and Miller, R.D. 1956. Physical theory for capillary flow phenomena. *J. Appl. Phys.* 27: 324-332.
10. Nemes, A., Timlin, D.J., Pachepsky, Y.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.
11. 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.
12. Nosrati Karizak, F., Movahedi Naeni, S.A., and Hezarjaribi, A. 2012. Using Artificial Neural Networks to estimate saturated hydraulic conductivity from easily available soil properties. *J. Soil Manage. Sust. Prod.* 2: 1. 95-116. (In Persian)

13. Rasoulzadeh, P., Razavi-Galajogh, S., and Neishabouri, M.R. 2012. Evaluate the accuracy methods to estimate the saturated hydraulic conductivity of soil water. *J. Water Res. Agric.* 26: 3. 303-316. (In Persian)
14. Reynolds, W.D., and Elrick, D.E. 1987. Laboratory and numerical assessment of the Guelph permeameter method. *Soil Sci.* 144: 244-282.
15. Roshani, Gh.A., and Dehghani, A.A., and Bouma, J. 1989. Using soil survey data for quantitative land evaluation. *Advanced Soil Science.* 9: 177-213. (In Persian)
16. Schaap, M., and Leij, F. 1998. Using neural networks to predict soil water retention and soil hydraulic conductivity. *Soil and Till Res.* 47: 37-42.
17. 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.
18. 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.
19. 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.
20. 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.
21. Wösten, J.H.M. 1997. Pedotransfer functions to evaluate soil quality, P 221-245. In: E.G. Gregorich and M.R. Carter (Eds.), *Soil quality for crop production and ecosystem health.* Developments in Soil Science, Elsevier, Amsterdam.
22. Zeleke, T.B., and Si, B.C. 2005. Scaling Relationships between Saturated Hydraulic Conductivity and Soil Physical Properties. *Soil Sci. Soc. Amer. J.* 69: 1691-1702.
23. Zolfaghari, A.A., Soltani, M.T., and Afshary Sarmadian, T.F. 2012. Compare K-nearest neighbor methods and artificial neural networks to estimate the soil cation exchange capacity. *Soil Manage. Sust. Prod. J.* 3: 1. 77-94. (In Persian)



Comparing nonparametric k-nearest neighbor technique with ANN model for predicting soil saturated hydraulic conductivity

*A. Khashei Siuki¹, V.R. Jalali Moakhar², A.M. Noferesti³ and Y. Ramazani¹

¹Assistant Prof., Dept. of Water Engineering, University of Birjand,

²Assistant Prof., Dept. of Soil Science, Shahid Bahonar University of Kerman,

³Instructor, Dept. of Water Engineering, University of Birjand

Received: 12/30/2013; Accepted: 09/28/2014

Abstract

Background and Objectives: Soil saturated hydraulic conductivity is the most important physical parameter, but its measurement often is difficult because of practical and/or cost-related reasons. In this research, expert system 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 (K_s) from other easily available soil properties.

Materials and Methods: In this research 151 soil samples were collected from farmlands around Bojnourd and saturated hydraulic conductivity (K_s) was estimated from other soil properties including soil textural fractions, EC, pH, SP, OC, TNV, ρ_s and ρ_b . To evaluate artificial neural network systems all the data were divided into 3 parts, 50% for training, 25% were allocated for the test and other data for validation. Design of MLP models (Multilayer Perceptron) was performed by sigmoid functions and hidden layer. ANN for all models was selected with Levenberg-Marcoardet algorithm to training as a hidden layer threshold logsig function for hidden layer and tansig for output layer.

Results: Results showed that the accuracy of the parameter estimation, using parametric method of artificial neural network to compare with k-nearest neighbors for terms of all the parameters (with $r=0.97$, $EF=0.946$, $RMSE=8.798$, $ME=28.446$ and $CRM= -0.144$) compared to other methods input models was acceptable.

Conclusion: The results showed that different techniques have estimated, hydraulic conductivity coefficient partially. The non-parametric method k-nearest neighbor, focus on the estimation of the regression line 1: 1 has been studied more than the other methods. The best result was for artificial neural network method with all information bank. Performance index k-nearest neighbor method ($EF=57$ to 71%), is powerful indication of this technique for estimation of soil hydraulic conductivity based on other easily available parameters, including particle size distribution, soil saturation extract, electrical conductivity (EC), saturation percentage of soil (SP), the organic carbon (OC), total of neutralizing value (TNV), bulk and apparent specific density of soil. ANN method can be used to estimate saturated hydraulic conductivity especially when new data set available.

Keywords: Artificial Neural Network, K-nearest neighbor technique, Soil saturated hydraulic conductivity

* Corresponding Authors; Email: abbaskhashei@birjand.ac.ir

