نشریه آب و خاک
جلد ۲۴، شماره ۴، مهر - آبان ۱۳۸۹، ص. ۸۱۴-۸۰۴

Journal of Water and Soil
Vol. 24, No. 4, Sep-Oct 2010, p. 804-814

برآورد رطوبت در نقطه پژمردگی دائم و ظرفیت زراعی خاک‌های شمال و شمال شرق ایران با استفاده از روش‌های نزدیک‌ترین K همسایه و شبکه‌های عصبی مصنوعی

امیر حق وردی[1*]- بیژن قهرمان[2]- علی اصغر خشنود یزدی[3]- زهرا عربی[4]

تاریخ دریافت:۸۹/۱/۲۲
تاریخ پذیرش:۸۹/۵/۲۴

چکیده

ظرفیت زراعی و پژمردگی دائم مهمترین نقاط پتانسیلی در مدل‌سازی و مدیریت آب مورد نیاز محصولات کشاورزی می‌باشند. روش‌های مستقیم تعیین میزان رطوبت هزینه‌بر و گران می‌باشد. بنابراین استفاده از توابع انتقالی برای تبدیل خصوصیات زودیافت خاک به خصوصیات هیدرولیکی یک راهکار مناسب برای حل این مشکل است. در این پژوهش کارایی مدل‌های شبکه عصبی مصنوعی (NNs) آموزش داده شده با نمونه‌های خاک منتج از جریان خروجی چند مرحله‌ای (NeuroMultistep outflow) و مدل‌های نزدیک‌ترین K همسایه (KNN) در اشتقاق توابع انتقالی به منظور تعیین میزان رطوبت در ظرفیت زراعی و پژمردگی دائم برای ۱۲۲ نمونه خاک از شمال و شمال شرق ایران مورد بررسی قرار گرفت. همچنین تاثیر عوامل ورودی مختلف و نوع داده به کار رفته برای اشتقاق هر دو روش معین شد. نتایج حاصله نشان دادند که در کل روش KNN (RMSE= ۰/۰۲۷) نسبت به NNs (RMSE= ۰/۰۳۷) نتایج بهتری داشت. همچنین می‌توان گفت که حساسیت مدل‌های شبکه عصبی به کیفیت و نوع داده‌های به کار رفته برای آموزش بسیار بالاست و همگن نبودن داده‌ها باعث کاهش کارایی مدل‌های شبکه عصبی و افزایش ۱۰۰ درصدی خطا می‌شود. همچنین نتایج نشان دادند که در نظر گرفتن خصوصیات هیدرولیکی به عنوان متغیرهای ورودی در شبکه عصبی باعث ارتقاء نتایج مدل‌سازی می‌شود.

واژه های کلیدی: نزدیک ترین K همسایه، شبکه‌های عصبی مصنوعی، توابع انتقالی، ظرفیت زراعی، پژمردگی دائم

مقدمه

تعیین میزان رطوبت خاک در نقاط ظرفیت زراعی (FC) و پژمردگی دائم (PWP) برای مقاصد کشاورزی بسیار حیاتی می‌باشد. آگاهی از میزان رطوبت در این نقاط برای محاسبه عمق مورد نیاز آب آبیاری و تخمین آب موجود در زمین ضروری است. اگر ناحیه تحت مطالعه نسبتاً کوچک باشد و یا اینکه به اندازه کافی از لحاظ خصوصیات فیزیک خاک همگن باشد، می‌توان با انجام تعداد قابل قبولی از آزمایشات مزرعه‌ای تقریب‌های خوبی از میزان رطوبت در این نقاط به عمل آورد. اما اگر ناحیه مورد مطالعه نسبتاً وسیع باشد به طوری که تغییرات مکانی قابل توجهی در خصوصیات هیدرولیکی خاک مشاهده گردد تقریباً غیر ممکن است بتوانیم تعداد آزمایشات را به گونه‌ای تنظیم کنیم که علاوه بر این که تقریب خوبی از میزان رطوبت در نقاط مذکور به‌دست می‌آوریم، محدودیت‌های مادی و زمانی انجام پروژه را نیز رعایت کرده باشیم (۱۷) .

روش‌های غیر مستقیم در حقیقت به عنوان راه حلی برای مشکلات فوق به دنبال پیش‌بینی خصوصیات هیدرولیکی خاک با استفاده از عوامل زود یافت اندازه‌گیری شده خاک می‌باشند. مفهوم توابع انتقالی نخستین بار توسط بوما (۳) مطرح شد. توابع انتقالی خصوصیات خاصی از خاک را (عمدتاً خصوصیات هیدورلیکی) با استفاده از خصوصیات پایه خاک برآورد می‌کنند (۶). گام معمول در اشتقاق توابع انتقالی ارائه روابطی تجربی میان خصوصیات پایه خاک و عواملی است که بایستی پیش‌بینی شوند. روش‌های رگرسیونی(۱۳و ۲۱) و به تازگی شبکه‌های عصبی مصنوعی (۷، ۸، ۱۲، ۱۴ و ۱۶) دو ابزار پرکاربرد برای تهیه توابع انتقالی می‌باشند. ویژگی معمول توابع انتقالی امروزی این است که همگی آن‌ها بر پایه روش پارامتریک می‌باشند، به عبارت دیگر آن‌ها معادلاتی هستند که پارامترهای آن‌ها با برازش بر داده‌ها حاصل می‌شوند. بنابراین دارای چندین نقطه ضعف

۱و۲- دانشجوی دکتری و استاد گروه آبیاری و زهکشی، دانشگاه فردوسی مشهد
(*-نویسنده مسئول : Email:amirhaghverdii@gmail.com)
۳- مربی آموزشکده شیروان، دانشگاه فردوسی مشهد
۴-کارشناسی ارشد خاکشناسی، دانشگاه آزاد اسلامی واحد علوم و تحقیقات تهران

می‌باشند به طوری که شناختن معادله صحیح دشوار است و توزیع‌های احتمالاتی خطا در فضای داده‌ها همیشه یکسان نمی‌باشد. در حالتی که تعداد داده‌ها کم است نتایج برآورد می‌تواند بسیار اریب باشد. عموماً این معادلات نیاز به بازنگری و اشتقاق مجدد دارند، داده‌های جدید برای تهیه آن‌ها بایستی موجود باشند، و کاربر قادر نخواهد بود که به آسانی مجموعه داده اضافی را برای بهبود عملکرد در محدوده خاصی از خصوصیات خاک در معادلات پیشین در نظر بگیرد (۱۰).

شبکه‌های عصبی مصنوعی (NNs) ماحصل تلاش موازی و مرتبط چندین رشته علمی از جمله ریاضیات، فیزیک و عصب شناسی و کامپیوتر می‌باشد. به‌ندرت علم در یک زمینه خاص دچار چنین چالشی شده است، به‌طوری که امروزه استفاده از این مدل‌ها در زمینه‌های بسیار زیادی صورت گرفته و محبوبیتشان نیز روز به روز رو به افزایش است. تقریباً نتیجه تمامی پژوهش‌های پیشین نشان می دهد که شبکه‌های عصبی حداقل به اندازه مابقی روش‌های مورد استفاده برای اشتقاق توابع انتقالی خوب عمل می‌کنند و به خوبی بر مفروضات آماری درگیر با توابع انتقالی فائق می‌آیند. شاپ و لجی (۱۵) نشان دادند که کیفیت پیش‌بینی توابع انتقالی توسط شبکه‌های عصبی وابسته به مجموعه داده‌هایی است که برای آموزش آن‌ها استفاده می شود زیرا دقت پیش‌بینی مستقیماً متاثر از کیفیت و نوع داده‌ها می‌باشد. همچنین نمس و همکاران (۱۱) عنوان نمودند که اشتقاق توابع انتقالی با استفاده از داده‌های کم ولی مناسب بهتر از استفاده از پایگاه داده‌های بزرگتر ولی عمومی و نامناسب می‌باشد. بنابراین دقت پیش‌بینی‌ها هنگامی که داده‌های هیدرولیکی و پایه خاک مورد استفاده برای آموزش با روش‌هایی مشابه به‌دست آمده باشند افزایش می‌یابد (۹). لذا سنجش مدل‌هایی با عوامل ورودی مختلف و برآورد تاثیر نوع داده‌های اشتقاق و آزمون بر کارایی روش های مختلف موضوع بسیاری از تحقیقات انجام شده در سالیان اخیر می‌باشد. میناسنی و همکاران (۹) تکنیک مدل‌سازی معکوس جریان خروجی چند مرحله‌ای[1] را به عنوان راه‌کاری جهت همگن‌سازی نمونه‌های خاک آموزش شبکه‌های عصبی به کار بردند. نتایج تحقیق آن‌ها حاکی از بهبود مدل‌های شبکه عصبی در تلفیق با جریان خروجی چند مرحله‌ای بود. به عبارت دیگر اگر داده‌های هیدرولیکی خاک با استفاده از یک روش واحد که برای تمامی نمونه های خاک به صورت یکسان اجرا شده باشد بدست آیند، نتایج پیش‌بینی منحنی نگهداشت رطوبتی و هدایت هیدرولیکی غیر اشباع با استفاده از خصوصیات زودیافت خاک می‌تواند بهبود یابد. آن‌ها عنوان نمودند که کارایی مدل شبکه عصبی ارائه شده برای مجموعه داده‌های مستقل بایستی مورد آزمون قرار گیرد، لذا یکی از اهداف این مقاله سنجش کارایی مدل مذکور برای نمونه‌های خاک ایران می‌باشد.

یک روش جایگزین برای روش‌های پارامتریک جهت برآورد خصوصیات هیدرولیکی خاک، استفاده از روش‌های ناپارامتریک می‌باشد. چنین روش‌هایی بر پایه شناسایی طرح و استفاده از شباهت‌ها به جای برازش معادلات به داده‌ها می‌باشند. هنگامی که شکل روابط میان داده‌های ورودی و خروجی از قبل مشخص نمی‌باشد استفاده از الگوریتم ناپارامتریک مفید است (۴و۲۲). بنابراین برای خصوصیات هیدرولیکی خاک که چگونگی وابستگی آن‌ها به خصوصیات پایه خاک از پیش معلوم نمی‌باشد استفاده از این روش‌ها می تواند راه‌گشا باشد (۱۰). یکی از الگوریتم‌های ناپارامتریک، الگوریتم نزدیک‌ترین K همسایه[2] (KNN) است. نحوه عملکرد این الگوریتم بر اساس مشاهدات و نمونه‌ها می‌باشد، به عبارت دیگر KNN یک الگوریتم مشاهده و یادگیری از طریق نظارت و بازرسی است که هدف آن دسته‌بندی یک عضو یا شیئ جدید بر اساس ویژگی نمونه‌های آموزش دهنده می‌باشد. بنابراین می‌توان یک نمونه جدید را بر اساس اکثریت K گروه و دسته که نزدیک‌ترین همسایگی‌ها را با آن نمونه داشته باشند، تقسیم بندی نمود. به طورکلی می‌توان گفت روش مذکور K تعداد از الگوهای مشابه را پیدا کرده و بر اساس آن‌ها ارزش نمونه مورد مطالعه را پیش‌بینی می‌کند. نمس و همکاران (۱۰) روش KNN را برای برآورد نقاط نگهداشت آب استفاده کردند و عملکرد آن‌ها را با مدل‌های شبکه عصبی مقایسه نمودند. نتایج تحقیق آن‌ها با معیار قرار دادن آماره‌های ریشه مجموع مربعات باقی مانده‌ها و باقی مانده‌های متوسط نشان داد که روش KNN دارای عملکردی به خوبی NNs می‌باشد. به نظر می‌رسد که این روش می‌تواند گزینه‌ای رقابتی برای دیگر روش‌های برآورد توابع انتقالی باشد. از آن جا که اساس این روش بر مبنای شباهت‌های موجود بین داده‌های اشتقاق و آزمون است سنجش کارایی آن برای مجموعه داده‌های مختلف در سرتاسر دنیا امری بسیار ضروری می‌باشد.

در شرایطی که نمونه‌های خاک مورد استفاده برای اشتقاق این مدل‌ها (KNN و NNeMultistep) نمونه‌های خاک دیگر مناطق دنیا باشد، سوال اساسی که بایستی به آن پاسخ داده شود میزان کارائی روش‌های مذکور برای نمونه خاک‌های ایران است و دیگر اینکه اگر در شرایط کمبود داده تصمیم به اشتقاق توابع انتقالی بر اساس اطلاعات محلی شود کدام یک از روش‌های فوق عملکرد مطلوب‌تری خواهند داشت. لذا یکی از اهداف فرعی این پژوهش تحلیل نحوه تاثیر نوع داده‌های مورد استفاده برای اشتقاق توابع انتقالی (بانک داده یا داده‌های محلی) در هر دو روش KNN و NNeMultistep می‌باشد. با توجه به تفاوت ماهوی نمونه خاک‌های مناطق مختلف دنیا، عدم استفاده از روش‌های فوق الذکر برای خاک

1-Multistep Outflow

2- k-Nearest Neighbor

های ایران تاکنون و خصوصیات متفاوت و منحصر به فرد هر کدام از آن‌ها، هدف از تحقیق حاضر توسعه و مقایسه روش‌های KNN و NNeMultistep به منظور برآورد میزان رطوبت در نقاط ظرفیت زراعی و پژمردگی دائم با استفاده از نمونه‌های خاک مناطقی از شمال و شمال شرق ایران می‌باشد. علاوه بر این با توجه به اینکه مرور منابع نشان می‌دهد در نظر گرفتن عوامل ورودی مختلف موجب تغییر در نتایج در هر دو روش KNN و NNs می‌شود (۹و۱۰)، دیگر هدف فرعی بررسی تاثیر نوع و تعداد عوامل ورودی بر عملکرد توابع انتقالی می‌باشد.

مواد و روش‌ها

نمونه‌های خاک مورد استفاده

یکصد و بیست و دو نمونه خاک برای انجام این تحقیق از مناطق آمل و بابل (۱۸ نمونه)، کرج (۳۲ نمونه) و خراسان (۷۲ نمونه) انتخاب شد (۱و۲). جدول شمارهٔ ۱ مشخصات فیزیکی خاک‌های مورد مطالعه را به صورت آماری نشان می‌دهد. از اراضی زراعی مختلف مناطق آمل، بابل وکرج ۵۰ نمونه خاک دست نخورده و دست خورده سطحی(۳۰-۰ سانتی متر) انتخاب شد و به همین علت کربن آلی آنها نسبت به خاک‌های مرتعی و جنگلی شمال کمتر است. نمونه‌برداری‌ها در نیمه دوم فصل بهار و به صورت چهارگوشه بندی معمولی و با فواصل یکسان انجام گرفت. نمونه‌های دست خورده خاک پس از خشک شدن در هوا و عبور از الک ۲ میلی‌متری برای انجام آزمایش‌ها آماده شد. مادهٔ آلی به روش اکسایشتر و بافت خاک به روش هیدرومتری پس از حذف آهک و مواد آلی تعیین شد (۲۰). مقادیر رطوبت وزنی خاک‌های دست نخورده در پتانسیل‌های ماتریک۱۵۰۰- ،۵۰۰-، ۱۰۰-،۳۳- و ۵- کیلوپاسکال با استفاده از صفحات فشاری در سه تکرار اندازه‌گیری شد. جرم مخصوص حقیقی کلیهٔ خاک‌ها ۲/۶۵ گرم بر سانتی‌متر مکعب در نظر گرفته شد.

علاوه بر این ۷۲ سری داده نیز از شمال شرق کشور مورد استفاده قرار گرفتند. نمونه برداری‌ها به صورت زیگزاگ انجام شد و بافت خاک به روش هیدرومتری و میزان رطوبت با استفاده از دستگاه صفحات فشاری در سه تکرار و در مکش‌های ۱۵۰۰-، ۱۰۰۰-، ۷۰۰-، ۴۰۰-، ۱۰۰- و ۳۳- تعیین شد (۲). میزان کربن آلی برابر ۵۸ درصد کل مواد آلی در نظر گرفته شد (۱۸). جدول ۱ نشان می‌دهد که میزان کربن آلی از مقادیر کم تا نسبتاً زیاد متغیر است و پراکنش توزیع اندازه ذرات خاک در نمونه‌های بجنورد بسیار بیش‌تر از دو منطقه دیگر می‌باشد.

روش KNN

بر خلاف توابع انتقالی کلاسیک روش KNN از هیچ تابع ریاضی از پیش تعیین شده‌ای برای برآورد یک نمایه خاص استفاده نمی‌کند. داده‌های مورد نیاز برای اجرای این روش در این پژوهش به دو دسته داده‌های مرجع (یا همان تعداد K همسایه) و داده‌های کاربردی تقسیم شدند. داده‌های مرجع مشابه با داده‌های آموزش یا واسنجی در روش‌های کلاسیک می‌باشند.

جدول ۱- مشخصات فیزیکی خاک‌های مورد استفاده در این پژوهش (به نقل از مرجع ۲ و ۱)

	داده‌های کرج (n=۳۲)					داده‌های آمل و بابل (n=۱۸)					داده‌های خراسان (n=۷۲)				
عوامل*	Max	Min	Mean	SD	CV	Max	Min	Mean	SD	CV	Max	Min	Mean	SD	CV
شن	۵۶/۸	۲۲	۴۰/۲۲	۷/۹۷	۱۹/۸۱	۵۰	۱۴/۸	۳۴/۳۲	۹/۳۳	۲۷/۲۱	۷۷/۲	۱۰/۲	۳۹/۲	۱۵/۱	۳۸/۴۵
سیلت	۵۰	۲۷/۴	۳۳/۹۵	۵/۱۳	۱۵/۱۲	۵۲	۲۷/۲	۳۴/۷۴	۶/۱۳	۱۷/۶۴	۵۹/۶	۱۴/۶	۴۲/۱	۱۱	۲۶/۱
رس	۴۰/۸	۱۴	۲۵/۸۹	۶/۵۹	۲۵/۴۶	۵۶	۱۸	۳۰/۳۳	۸/۶۸	۲۸/۶۳	۳۲/۲	۸/۲	۱۸/۷	۵/۹	۳۱/۸۱
کربن آلی	۰/۹۹	۰/۲۰	۰/۶۱	۰/۲۲	۳۶/۶	۱/۸۴	۰/۶۸	۱/۲۴	۰/۳۳	۲۶/۹	۱/۴	۰/۱۲	۰/۶	۰/۲۷	۴۴/۳۴
BD	۱/۶۳	۱/۳۷	۱/۴۷	۰/۰۵۷	۳/۹	۱/۵۳	۱/۳۹	۱/۴۷	۰/۰۴	۲/۶۹	۱/۵۱	۱/۳۷	۱/۴۲	۰/۰۳	۲/۰۲
θ_{**}	۶۶/۳	۴۶/۶	۵۴/۶۲	۵/۰۷	۹./۲۹	۷۸/۵	۵۴/۹	۶۲/۲۴	۶/۴۵	۱۰/۳۴	–	–	–	–	–
$\theta_{-۵}$	۴۹/۱	۳۳	۴۰/۷	۳/۸۷	۹/۵۱	۵۵/۳	۳۸/۱	۴۳/۹۷	۴/۲۵	۹/۷۲	–	–	–	–	–
$\theta_{-۳۳}$	۳۵/۷	۲۰	۲۹/۵۲	۳/۶۱	۱۲/۲۳	۳۹/۷	۲۵/۲	۳۳/۱۵	۴/۱۴	۱۲/۲۴	۳۸/۱	۲۲/۲	۳۰/۳۳	۴/۵۹	۱۵/۱۵
$\theta_{-۱۰۰}$	۳۱/۳	۱۶/۳	۲۳/۸۱	۳/۱۰	۱۳/۰۴	۳۵/۱	۲۲/۳	۲۷/۴۲	۳/۶۹	۱۳/۴۱	۳۶/۲۳	۱۹/۲	۲۶/۵۵	۴/۱۳	۱۵/۵۶
$\theta_{-۴۰۰}$	–	–	–	–	–		–	–	–	–	۲۷/۸۲	۱۶	۲۲/۱۹	۲/۷۹	۱۲/۵۵
$\theta_{-۵۰۰}$	۲۳/۹	۱۱/۵	۱۸/۰۵	۲/۷	۱۴/۹۸	۲۷/۶	۱۵/۲	۲۰/۸۲	۳/۰۱	۱۴/۱۳	–	–	–	–	–
$\theta_{-۷۰۰}$	–	–	–	–	–		–	–	–	–	۲۳/۰۸	۱۱/۸	۱۷/۲۱	۳/۰۳	۱۷/۵۹
$\theta_{-۱۰۰۰}$	–	–	–	–	–		–	–	–	–	۲۱/۵۴	۹/۵۵	۱۵/۶۹	۳	۱۹/۱۲
$\theta_{-۱۵۰۰}$	۳۰/۹	۹/۳	۱۵/۴۳	۴/۱۲	۲۶/۶۸	۲۶/۹	۱۴/۸	۱۸/۵۵	۳/۰۳	۱۵/۹۴	۲۰/۰۸	۸/۳۸	۱۴/۶۴	۲/۹	۱۹/۸۴

* CV ضریب تغییرات، SD انحراف استاندارد، Max بیشترین، Min کمترین، Mean میانگین و n تعداد نمونه‌های خاک می‌باشند.

** منظور از θ_{-p} میزان رطوبت حجمی خاک در پتانسیل p- کیلو پاسکال بر حسب درصد و BD جرم مخصوص ظاهری خاک بر حسب گرم بر سانتی متر مکعب می‌باشد.

در بیش‌تر پژوهش‌های انجام شده با استفاده از روش KNN "فاصله" به معنی مفهوم رایج اقلیدسی آن بین هدف و نمونه‌های معلوم می‌باشد. فاصله هر نمونه خاک از خاک هدف می‌تواند با استفاده از تفاوت‌های ریشه مجموع مربعات خطا در عوامل ورودی میان خاک هدف و هر کدام از نمونه‌های خاک مرجع مشخص شود. سپس خاک‌های مرجع به صورت صعودی بر اساس فاصله‌شان تا نمونه خاک مذکور مرتب می‌شوند. مقدار تقریب زده شده نمایه خروجی (در این تحقیق رطوبت در نقاط ظرفیت زراعی و پژمردگی دائم) به عنوان متوسط وزنی نمایه‌های خروجی نزدیک‌ترین خاک‌های انتخابی می‌باشد. غالباً برای جلوگیری از ایجاد اریب در جهت هر کدام از عوامل ورودی، داده‌های مورد استفاده نرمال می شوند (۱۰).

از آن جا که یکی از اهداف فرعی این پژوهش عملکرد بررسی تاثیر نوع بانک داده مورد استفاده برای اشتقاق توابع انتقالی بود دو رویه کلی جهت تخصیص داده‌ها و اجرای روش KNN استفاده شد. در حالت اول از بانک داده US NRCS-SCS (۱۹) به عنوان داده‌های مرجع استفاده شد. دلیل انتخاب این بانک داده آن بود که این بانک داده توسط نمس و همکاران (۱۰) نیز استفاده شده بود و امکان مقایسه بهتر نتایج این پژوهش با تحقیق آن‌ها را فراهم می کرد. در این حالت داده‌های ایران به تفکیک در هر سه منطقه به عنوان مجموعه داده‌های کاربردی به کار رفتند. در حالت دوم تنها از داده‌های ایران به عنوان داده‌های مرجع و همچنین داده‌های کاربردی استفاده شد. برای اعتبارسنجی مدل‌ها از روش اعتبارسنجی چند-قسمتی[1] استفاده شد (۵). بدین منظور نمونه‌های خاک ایران به ۱۰ قسمت مساوی تقسیم شدند (۸ قسمت ۱۲ نمونه‌ای و دو قسمت ۱۳ نمونه‌ای) و در هر بار ۹ قسمت به عنوان داده‌های مرجع و ۱ قسمت باقیمانده به عنوان داده‌های کاربردی مورد استفاده قرار گرفتند. در نهایت نتایج برای کلیه داده‌ها و همچنین نتایج تفکیکی بر اساس مناطق سه‌گانه مورد بررسی قرار گرفتند. دیگر هدف فرعی این پژوهش تحلیل نحوه تاثیر نوع و تعداد عوامل ورودی بر عملکرد روش‌های مورد استفاده بود. بنابراین چهار مدل موجود در جدول ۲ جهت تعیین تاثیر نوع و تعداد متغیرهای ورودی بر عملکرد روش مورد استفاده قرار گرفتند. تمامی محاسبات مربوط به روش KNN توسط نرم افزار k-Nearest Neighbor (http://www.ars.usda.gov) انجام شد.

روش Multistep Outflow و مدل‌های NNs[2]

روش جریان خروجی چند مرحله‌ای، مدل‌سازی معکوس را برای برآورد منحنی هدایت هیدرولیکی و نگهداشت آب خاک در یک آزمایش واحد ناپایدار وابسته به زمان به کار می‌بندد. این روش یکی از روش‌های رایج برای برآورد خصوصیات هیدرولیکی خاک می‌باشد (۹). روش معکوس ترکیبی از آزمایشات با مدل‌سازی عددی را به کار می‌گیرد، بنابراین هم به یک روش آزمایشی دقیق و هم به مدل‌سازی عددی و الگوریتم‌های بهینه‌سازی احتیاج دارد (۹). در واقع هدف از تلفیق این روش و مدل‌های شبکه عصبی توسط میناسنی و همکاران (۹)، فراهم کردن نمونه‌های خاکی بوده است که علاوه بر تشابه روش به‌دست آوردن خصوصیات هیدرولیکی، تمامی دامنه تغییرات خصوصیات پایه خاک را پوشش دهد.

جدول ۲- مدل‌های در نظر گرفته شده برای روش KNN

مدل	ورودی‌های مورد استفاده
KNN_1	Sand, Silt, Clay
KNN_2	Sand, Silt, Clay, BD
KNN_3	Sand, Silt, Clay, OM
KNN_4	*Sand, Silt, Clay, BD, OM

* sand، silt و clay به ترتیب درصدهای شن، سیلت و رس و BD و OM جرم مخصوص ظاهری و میزان ماده آلی هستند.

با در نظر گرفتن مطالب بالا، برای اجرای مدل‌های شبکه عصبی از نرم افزار NeuroMultistep (www.usyd.edu.au) استفاده شد. وزن‌های مدل شبکه عصبی مذکور مذکور توسط میناسنی و همکاران (۹) ارائه شده است. مدل شبکه عصبی ارائه شده پیش‌خور با تابع انتقال سیگموئید می‌باشد. آن‌ها با استفاده از ۳۱۰ نمونه خاک از ۳ منطقه مختلف که خصوصیات هیدرولیکی تمامی آن‌ها با استفاده از روش خروجی چند مرحله‌ای به‌دست آمده بود، ساختار بهینه و وزن‌های این مدل را ارائه کردند. ترکیبات عوامل ورودی ذکر شده در جدول ۳ مورد آزمایش قرار گرفتند. لازم به ذکر است که به علت اینکه نمونه‌های خاک بجنورد فاقد میزان رطوبت اندازه‌گیری شده اشباع و هدایت هیدرولیکی اشباع بودند فقط مدل‌های ۱ و ۲ جدول ۳ برای آن‌ها مورد بررسی قرار گرفت. از تکنیک اعتبار سنجی چند قسمتی شرح داده شده برای درنظر گرفتن تمامی نمونه‌های خاک در فاز آزمون استفاده شد. برای بررسی حالتی که صرفاً وزن‌های مدل‌های شبکه عصبی با استفاده از نمونه‌های ایران به‌دست آیند از نرم افزار Neorusolution5.07 (www.nd.com) استفاده شد. شبکه پرسپترون چند لایه، الگوریتم آموزش لونبرگ مارکوات و تابع انتقال در لایه میانی تانژانت هیپربولیک و در لایه خروجی خطی بود. تعداد نرون‌های لایه میانی از ۱ تا ۱۵ تغییر داده شد. در این حالت نیز تمامی مدل‌های موجود در جدول ۳ مورد بررسی قرار گرفتند. توجه شود که با توجه به این امر که Ks خود جزء خصوصیات دیریافت

1- N-fold validation
2 -Neural Networks

خاک می‌باشد، در نظر گرفتن این نمایه در ورودی مدل NNs_4 صرفاً کوششی تئوریک در جهت آزمودن میزان ارتقاء نتایج مدل سازی منحنی رطوبتی توسط مدل‌های شبکه عصبی می‌باشد.

جدول ۳- مدل‌های در نظر گرفته شده برای روش NNs

مدل	ورودی‌های مورد استفاده
NNs_1	Sand, Silt, Clay
NNs_2	Sand, Silt, Clay, BD
NNs_3	Sand, Silt, Clay, BD, θ_s
NNs_4	Sand, Silt, Clay, BD, θ_s, K_s

* K_s و θ_s به ترتیب هدایت هیدرولیکی اشباع و رطوبت اشباع و مابقی نمادها مشابه جدول ۲ می‌باشند.

معیارهای ارزیابی

به منظور سنجش کارایی شبکه‌های مختلف مورد بحث در برآورد نمایه‌های مختلف خروجی مدل‌سازی، از آماره‌های ضریب همبستگی (معادله ۱)، ریشه میانگین مربعات خطا (معادله ۲) و میانگین مربعات خطای نرمال شده (معادله ۳) استفاده شد (۵).

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(M_i - E_i)}{\sum_{i=1}^{n} M_i^2 - \frac{\sum_{i=1}^{n} E_i^2}{n}} \quad (۱)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(E_i - M_i)^2}{n}} \quad (۲)$$

$$NMSE = \frac{\sum_{i=1}^{n}(E_i - M_i)^2}{\sum_{i=1}^{n} E_i M_i} \quad (۳)$$

که در آن‌ها M بیانگر مقادیر واقعی رطوبت، E مقادیر تخمینی و n تعداد نمونه‌های فاز آزمون می‌باشد.

نتایج و بحث

تحلیل چیدمان عوامل ورودی

مدل‌های KNN

نتایج حاصل از مدل‌سازی توسط مدل‌های KNN و NNs در جدول ۴ نشان داده شده است. در نمونه خاک‌های آمل- بابل بهترین نتایج به طور میانگین در نقطه پژمردگی دائم در مدل KNN_4 با آماره‌های RMSE، NMSE و r به ترتیب برابر ۰/۰۳۴، ۰/۰۳ و ۰/۹۲ به‌دست آمد، در حالی که نتایج مدل‌های مورد آزمون قرار گرفته در نقطه ظرفیت زراعی تفاوت فاحشی با یکدیگر نداشتند. مدل KNN_4 در کرج نیز با آماره‌های RMSE، NMSE و r به ترتیب برابر با ۰/۰۲۴، ۰/۰۰۷ و ۰/۷۴ در ظرفیت زراعی و با آماره‌های RMSE، NMSE و r به ترتیب برابر با ۰/۰۳۲، ۰/۰۳۶ و ۰/۸۷ در نقطه پژمردگی دائم بهترین نتایج را دارا بود. در داده‌های بجنورد اما نتایج مدل‌های مورد بررسی تفاوت چندانی با یکدیگر نداشتند. برتری نسبی نتایج مدل‌سازی در نقطه پژمردگی دائم نسبت به ظرفیت زراعی در تمامی مدل‌ها مشاهده شد. در مورد نمونه‌های خاک کرج البته این تفاوت زیاد چشم‌گیر نبود. ضعیف‌تر بودن نتایج در ظرفیت زراعی نسبت به پژمردگی دائم در مدل‌های KNN توسط نمس و همکاران (۱۰) نیز گزارش شده بود. احتمالاً دو دلیل ممکن است باعث این امر شود. اول اینکه مقادیر رطوبت در ظرفیت زراعی بیش‌تر از پژمردگی دائم می‌باشد و دوم اینکه میزان رطوبت در ظرفیت زراعی بیشتر از پژمردگی دائم تحت تاثیر ساختمان خاک می‌باشد و ساختمان خاک در مدل‌های بررسی شده تنها در عامل BD منعکس می‌شوند (۱۰). در نظر گرفتن عوامل BD و OM به عنوان ورودی در مدل‌های KNN تغییر قابل توجهی در نتایج نداشت. این امر نیز مشابه با تحقیق نمس و همکاران (۱۰) بود. آن‌ها عنوان نمودند که بهبود بسیار اندکی و از لحاظ آماری بی معنی در اثر استفاده از عوامل ورودی BD و OM به‌دست می‌آید. در کل می‌توان گفت که مدل‌های KNN عملکرد قابل قبولی از خود نشان دادند.

مدل‌های NNs

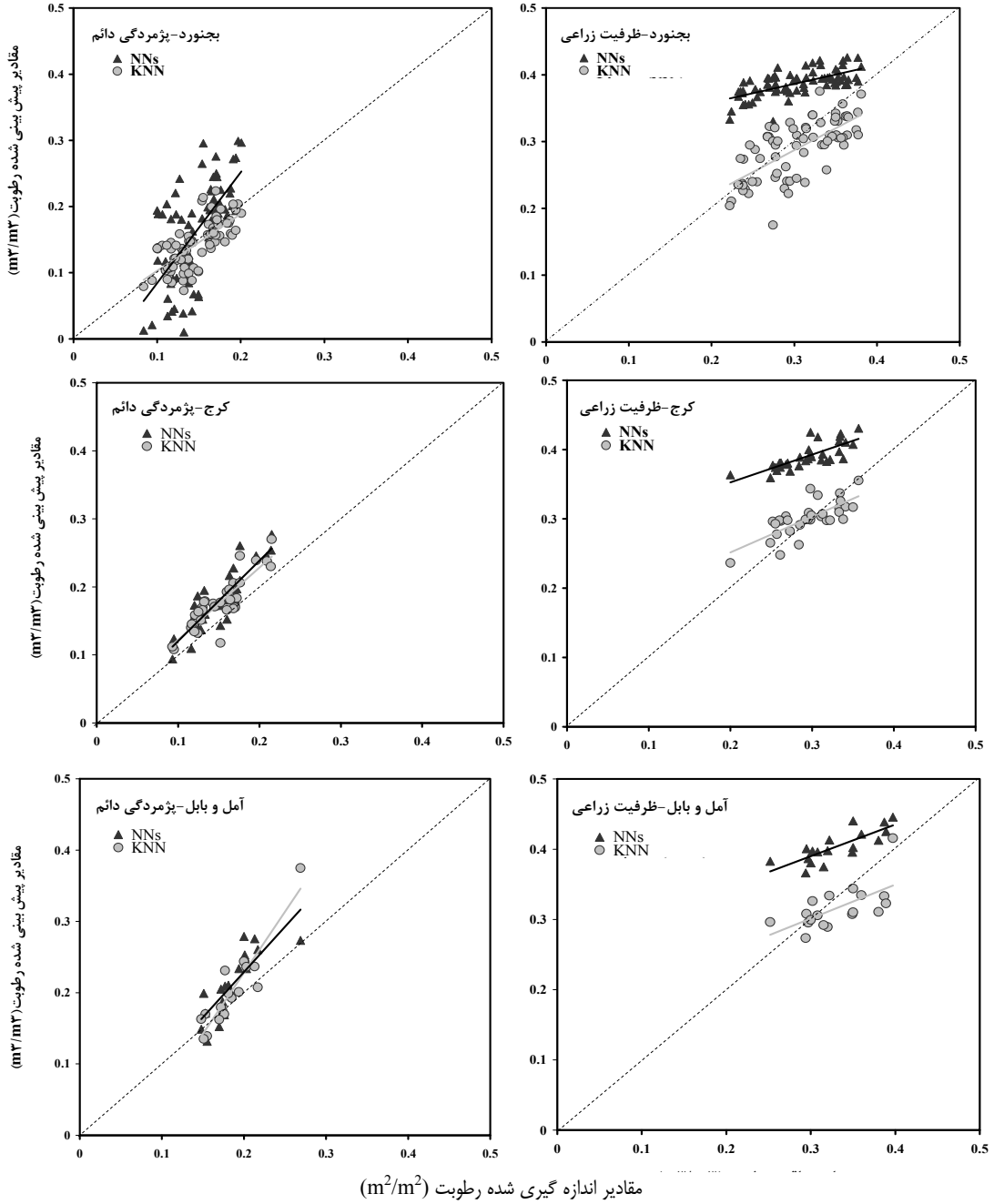
در خاک‌های آمل و بابل بهترین نتایج مدل‌سازی توسط NNs متعلق به NNs_4 با آماره‌های RMSE، NMSE و r در نقطه ظرفیت زراعی به ترتیب برابر ۰/۰۸، ۰/۰۴ و ۰/۸ و در پژمردگی دائم به ترتیب برابر ۰/۰۳۷، ۰/۰۳ و ۰/۸۲ بود. در نمونه خاک‌های کرج نیز به طریقی مشابه مدل NNs_4 با آماره‌های RMSE، NMSE و r در نقطه ظرفیت زراعی به ترتیب برابر ۰/۰۹۸، ۰/۰۸۴ و ۰/۷۷ و در پژمردگی دائم به ترتیب برابر ۰/۰۳۷، ۰/۰۴۹ و ۰/۸۶ بهترین نتایج را داشت. در بجنورد نتایج مدل‌های NNs_1 و NNs_2 در نقطه ظرفیت زراعی تقریباً مشابه بودند و در نقطه پژمردگی دائم مدل NNs_2 عملکرد مطلوب‌تری داشت. لذا با توجه به اعداد ذکر شده می‌توان گفت نتایج در آمل- بابل و کرج در هر دو نقطه پژمردگی دائم و ظرفیت زراعی از بجنورد بهتر بود. این تفاوت نتایج از آن جا ناشی می‌شد که به علت در اختیار نبودن عوامل θ_s و K_s برای نمونه خاک‌های بجنورد تنها مدل‌های NNs_1 و NNs_2 مورد بررسی قرار گرفته بودند. این امر و بررسی دقیق‌تر روند تغییرات در نتایج مدل‌های مختلف شبکه عصبی به خوبی نشان می‌دهد که با اضافه نمودن عوامل θ_s و K_s نتایج بهبود محسوسی پیدا کرده‌اند.

جدول ٤- نتایج مدل‌سازی توسط مدل‌های KNN و NNs

نوع مدل	$\theta_{۳۳}$ RMSE	$\theta_{۳۳}$ NMSE	$\theta_{۳۳}$ r	$\theta_{۱۵۰۰}$ RMSE	$\theta_{۱۵۰۰}$ NMSE	$\theta_{۱۵۰۰}$ r
آمل- بابل						
NNs_1	۰/۱۵۴	۰/۱۵	۰/۷	۰/۱۴	۰/۳۱	۰/۲۵
NNs_2	۰/۱۵۲	۰/۱۴	۰/۷۳	۰/۱۲	۲	۰/۷۷
NNs_3	۰/۰۸	۰/۰۵	۰/۷۱	۰/۰۶۲	۰/۰۸	۰/۸۶
NNs_4	۰/۰۸	۰/۰۴	۰/۸	۰/۰۳۷	۰/۰۳	۰/۸۲
KNN_1	۰/۰۳	۰/۰۰۹	۰/۶۳	۰/۰۴۱	۰/۰۴	۰/۹۳
KNN_2	۰/۰۳	۰/۰۰۹	۰/۶۳	۰/۰۴	۰/۰۴	۰/۹۲
KNN_3	۰/۰۳۴	۰/۰۱	۰/۶۴	۰/۰۳۷	۰/۰۳	۰/۹۱
KNN_4	۰/۰۳۴	۰/۰۱	۰/۶۳	۰/۰۳۴	۰/۰۳	۰/۹۲
کرج						
NNs_1	۰/۱۶	۰/۱۹	۰/۶۱	۰/۱۵	۰/۴۸	۰/۷۲
NNs_2	۰/۱۶	۰/۱۹	۰/۶۶	۰/۱	۰/۲۷	۰/۸۱
NNs_3	۰/۱	۰/۰۹	۰/۷۲	۰/۰۶۷	۰/۱۳	۰/۸۶
NNs_4	۰/۰۹۸	۰/۰۸۴	۰/۷۷	۰/۰۳۷	۰/۰۴۹	۰/۸۶
KNN_1	۰/۰۳۵	۰/۰۱	۰/۶۲	۰/۰۴۱	۰/۰۵۷	۰/۸۷
KNN_2	۰/۰۲۸	۰/۰۰۸	۰/۷۳	۰/۰۴	۰/۰۵۵	۰/۸۶
KNN_3	۰/۰۲۵	۰/۰۰۷	۰/۷۳	۰/۰۳۴	۰/۰۴	۰/۸۸
KNN_4	۰/۰۲۴	۰/۰۰۷	۰/۷۴	۰/۰۳۲	۰/۰۳۶	۰/۸۷
بجنورد						
NNs_1	۰/۰۹	۰/۰۶۹	۰/۵۸	۰/۰۵۲	۰/۱۱	۰/۶۲
NNs_2	۰/۰۹	۰/۰۷۱	۰/۶۵	۰/۰۶۲	۰/۱۵	۰/۶۴
KNN_1	۰/۰۳۸	۰/۰۱۶	۰/۶۹	۰/۰۲۶	۰/۰۳۱	۰/۷
KNN_2	۰/۰۴۱	۰/۰۱۹	۰/۶۹	۰/۰۲۵	۰/۰۲۹	۰/۷۳
KNN_3	۰/۰۴۴	۰/۰۲۳	۰/۶۸	۰/۰۲۷	۰/۰۳۵	۰/۷۲
KNN_4	۰/۰۴۶	۰/۰۲۵	۰/۶۹	۰/۰۲۶	۰/۰۳۳	۰/۷۲

این بهبود نتایج با افزایش عوامل ورودی در پژوهش میناسنی و همکاران (۹) نیز گزارش شده بود. روند تغییرات RMSE در پژوهش حاضر و پژوهش میناسنی و همکاران (۹) در میان مدل‌های در نظر گرفته شده مشترک مشابه می‌باشد. این تشابه نشان می‌دهد که در نظر گرفتن نمایه‌هایی هیدرولیکی همچون K_s به عنوان ورودی موجب می‌شود که شبکه‌های عصبی عملکرد بهتری را از خود نشان دهند. به عنوان مثال در تحقیق آن‌ها مقادیر آماره RMSE مدل‌های NNs_1 و NNs_4 به ترتیب برابر با ۰/۰۴۲ و ۰/۰۳۵ بود.

در تحقیق میناسنی و همکاران (۹) مقادیر آماره RMSE توسط شبکه‌های عصبی در کل کوچکتر از مقادیر به‌دست آمده در این پژوهش بود. این تفاوت به خوبی نشان می‌دهد که با وجود اینکه داده‌های فاز آزمون همگی با روش جریان خروجی چند مرحله‌ای توسعه یافته بودند، هنگامی که نمونه‌هایی خارج از مجموعه آزمایش و از یک منطقه دیگر توسط یک مدل شبکه عصبی مورد آزمون قرار می‌گیرند نتایج به نسبت میزان تفاوت ماهوی نمونه‌های آزمون و آزمایش ضعیف‌تر می‌شوند. بر طبق نتایج حاصله می‌توان چنین عنوان نمود که استفاده از روش جریان خروجی چند مرحله‌ای و در نتیجه همگنی داده‌های آموزش موجب دستیابی به عملکردی قابل قبول توسط شبکه‌های عصبی شده است. با توجه به اینکه در پژوهش‌های مرسوم انجام شده توسط مدل‌های شبکه عصبی نیاز به داده‌های بسیاری برای آموزش وجود دارد به نظر می‌رسد استفاده از بانک داده ای که خصوصیات هیدرولیکی آن با استفاده از یک روش واحد توسعه یافته باشد بهتر از استفاده از بانک داده‌ای است که از روش‌های متعددی برای تعیین خصوصیات هیدرولیکی خاک‌های آن استفاده شده باشد. استدلالی مشابه با مدل‌های KNN در ارتباط با نتایج ضعیف‌تر مدل‌سازی در ظرفیت زراعی نسبت به پژمردگی دائم در این جا نیز صادق می‌باشد.



شکل۱- پراکنش بهترین نتایج مدل‌سازی KNN و NNs به تفکیک مناطق در نقاط ظرفیت زراعی و پژمردگی دائم حول خط ۱:۱

مقایسه عملکرد دو روش

جهت مقایسه بهتر بهترین نتایج حاصله توسط مدل‌های KNN و NNs، پراکنش مقادیر پیش‌بینی شده رطوبت در نقاط ظرفیت زراعی و پژمردگی دائم حول خط ۱:۱ و همچنین میزان همبستگی در شکل ۱ نشان داده‌شده‌است. بررسی شکل ۱ نشان می‌دهد که در نقطه ظرفیت زراعی در تمامی سه سایت مدل‌های KNN پراکنش بهتری و همبستگی بیش‌تری نسبت به مدل‌های NNs دارند. نتایج این مدل‌ها تقریباً به طور قرینه در حول خط ۱:۱ توزیع شده‌اند و از طرف دیگر مدل‌های NNs در تمامی مناطق بیش برآورد دارند و در بالای خط ۱:۱ پراکنده شده‌اند. در نقطه پژمردگی دائم بهبود پراکنش نقاط نسبت به نقطه ظرفیت زراعی بسیار مشهود است. مدل‌های NNs و KNN در نقطه پژمردگی دائم تقریباً پراکنش یکسانی دارند هر چند

برتری نسبی در این نقطه نیز با مدل‌های KNN می‌باشد. این نتایج را می‌توان به گونه‌ای از جدول ۴ نیز استنباط کرد. به عبارت دیگر در نقطه پژمردگی دائم نتایج در هر دو مدل NNs و KNN تقریباً یکسان می‌باشد، اما در ظرفیت زراعی عملکرد مدل‌های KNN بهتر می‌باشد.

در تحقیق Nemes نتایج مدل‌سازی توسط مدل‌های NNs و KNN تقریباً مشابه بود و برتری نسبی مدل‌های NNs از لحاظ آماری معنی‌دار نبود. این تفاوت نتیجه می‌تواند به این دلیل باشد که در این پژوهش برای مدل‌های شبکه عصبی ما از وزن‌های به‌دست آمده در تحقیق میناسنی و همکاران (۹) استفاده کرده‌ایم در حالی که در پژوهش نمس و همکاران (۱۰) این چنین نبوده است و نمونه‌های آزمایش و آزمون در تحقیق آن‌ها از یک منبع بوده است. علاوه بر این شاید بتوان گفت که حساسیت مدل‌های KNN به تفاوت بین داده هایی که برای اشتقاق و آزمون به کار می‌رود کمتر از مدل‌های NNs می‌باشد. این امر برای مدل‌های KNN یک برتری عمده محسوب می‌شود.

ما مدل‌های KNN و NNs با عوامل ورودی یکسان (مثلاً مدل KNN_1 با مدل NNs_1 و مدل KNN_2 با مدل NNs_2) را مقایسه کردیم. نتایج حاکی از برتری مدل‌های KNN بود. به گونه‌ای که به عنوان مثال در مدل NNs_1 میانگین آماره‌های RMSE، NMSE و r در نقطه ظرفیت زراعی در هر سه منطقه به ترتیب برابر ۰/۱۳، ۰/۱۴ و ۰/۶۳ و در مدل KNN_1 برابر با ۰/۰۳، ۰/۰۱ و ۰/۶۵ می‌باشد. در مدل NNs_1 میانگین آماره‌های RMSE، NMSE و r در نقطه پژمردگی دائم در هر سه منطقه به ترتیب برابر ۰/۱۱، ۰/۳ و ۰/۵۳ و در مدل KNN_1 برابر با ۰/۰۴، ۰/۰۴ و ۰/۸۵ می‌باشد. بنابراین به نظر می‌رسد در حالتی که صرفاً از داده‌های اطلاعات پایه خاک استفاده شود و همچنین نمونه‌های خاک آموزش و مرجع با نمونه‌های فاز آزمون به هیچ عنوان تشابه منطقه‌ای ندارند مدل‌های KNN_1 نتایج بسیار مطلوب‌تری را نسبت به مدل‌های NNs_1 ارائه کردند. این یک نتیجه کاربردی بسیار ارزشمند محسوب می‌شود. زیرا در ایران داده کافی برای تحلیل ویژگی هیدرولیکی خاک وجود ندارد و استفاده از بانک داده جهانی برای بسیاری از مناطق یک ضرورت به شمار می‌آید. در پژوهش نمس و همکاران (۱۰) میانگین آماره RMSE برای مدل‌های KNN_1 و KNN_2 به ترتیب برابر ۰/۰۴۷ و ۰/۰۴۶ و برای مدل‌های NNs_1 و NNs_2 به ترتیب برابر ۰/۰۴۶ و ۰/۰۴۵ بود.

یعنی در شرایط استفاده از داده‌های مشابه پایه خاک عملکرد مدل‌های KNN و NNs تقریباً مشابه با یکدیگر بود. این تفاوت محتملاً به دلیل تفاوت ماهوی خاک‌های ایران با نمونه‌های خاک دیگر مناطق می‌باشد. سوالی که متعاقب این نتایج و تحلیل‌ها به ذهن می‌رسد این است که چرا این اختلاف نتایج در مدل‌های بهینه مناطق که عمدتاً مدل‌های KNN_4 و NNs_4 بودند (جدول ۴) با این شدت به چشم نخورد؟ به نظر می‌رسد منطقی‌ترین پاسخ مربوط به تفاوت در عوامل ورودی مدل‌های بهینه باشد. به عبارت دیگر عوامل اضافه شده به مدل‌های NNs، یعنی θ_s و K_s، نسبت به عوامل افزوده شده به مدل‌های KNN، یعنی BD و OM، تاثیر بیش‌تری در بهبود نتایج مدل سازی داشتند.

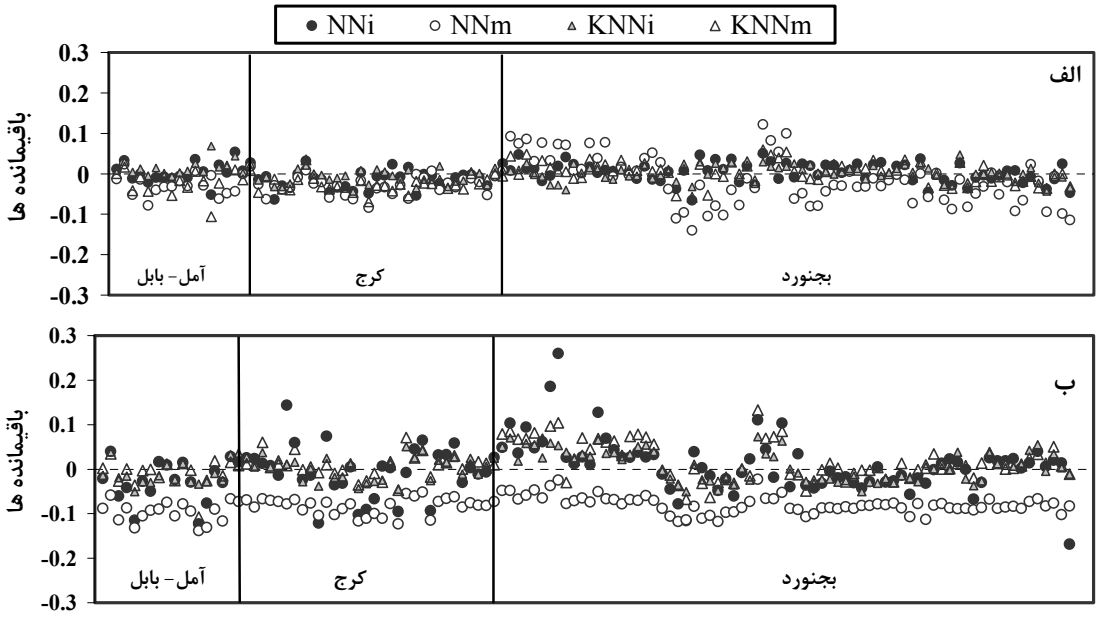
تحلیل نوع داده‌های آموزش و مرجع در عملکرد

نتایج کلی نمونه‌های خاک ایران در دو حالت استفاده از داده‌های ایران برای اشتقاق و آزمون (مشخص شده با اندیس i) و استفاده از بانک داده برای اشتقاق و داده‌های ایران برای آزمون (مشخص شده با اندیس m) در جدول ۵ نشان داده شده است. نتایج حاصله نشان می‌دهد که اگر از نمونه‌های خاک ایران به عنوان داده‌های آموزشی در مدل‌های شبکه عصبی استفاده شود باعث بهبود نسبی نتایج می‌شود. به گونه‌ای که میزان آماره RMSE در نقطه ظرفیت زراعی و پژمردگی دائم برای کل نمونه خاک‌های ایران در NN_i به ترتیب ۰/۰۵ و ۰/۰۲ و در NN_m به ترتیب ۰/۰۹ و ۰/۰۵ بود. در نتیجه استفاده از داده های ایران برای آموزش، در مقایسه با استفاده از داده های بانک، برای FC و PWP به ترتیب موجب ۵۳ و ۴۵ در صد کاهش در RMSE گردید. این مطلب در مورد مدل‌های KNN تنها در نقطه پژمردگی دائم صادق بود (۳۱ درصد کاهش) و در نقطه ظرفیت زراعی عملکرد دو مدل تقریباً یکسان (RMSE=۰/۰۳۴) بود. میزان آماره RMSE در نقطه ظرفیت زراعی و پژمردگی دائم برای کل خاک‌های ایران در KNN_i به ترتیب ۰/۰۳۴ و ۰/۰۲ و در KNN_m به ترتیب ۰/۰۳۴ و ۰/۰۲۹ بود. شاپ و لجی (۱۵) و نمس و همکاران (۱۱) نیز در تحقیقات جداگانه‌ای عنوان نمودند که در صورت موجود بودن نمونه خاک‌های همگن محلی اشتقاق توابع انتقالی با آن‌ها نتایج مطلوب‌تری از استفاده از داده‌های حجیم ولی متفاوت با داده‌های محلی دارد.

جدول ۵- نتایج کلی مدل‌سازی توسط مدل‌هایNNs و KNN به تفکیک نوع نمونه‌های مورد استفاده برای اشتقاق توابع

آماره	ظرفیت زراعی				پژمردگی دائم			
	NN_i*	NN_m	KNN_i	KNN_m	NN_i	NN_m	KNN_i	KNN_m
RMSE	۰/۰۵	۰/۰۹۱	۰/۰۳۴	۰/۰۳۴	۰/۰۲۵	۰/۰۵۳	۰/۰۲	۰/۰۲۹
NMSE	۰/۰۲۷	۰/۰۶۹	۰/۰۱۲	۰/۰۱۳	۰/۰۲۵	۰/۱	۰/۰۱۶	۰/۰۳۲
r	۰/۵۶	۰/۷	۰/۶۵	۰/۶۹	۰/۷۵	۰/۷	۰/۷۹	۰/۷۹

* NN_i و NN_m به ترتیب NNs اشتقاقی از داده‌های ایران و بانک داده و KNN_i و KNN_m به ترتیب مدل‌های KNN اشتقاقی از داده‌های ایران و بانک داده می‌باشند.



شکل ۲- پراکنش باقیمانده‌ها برای مدل‌های NN_i، NN_m، KNN_i و KNN_m. الف. پژمردگی دائم ب. ظرفیت زراعی

در شکل ۲ پراکنش باقیمانده‌ها (تفاوت مقادیر اندازه‌گیری شده (θ_m) و پیش‌بینی شده رطوبت (θ_p)؛ ($R= \theta_m-\theta_p$) برای مدل‌های NN_i، NN_m، KNN_i و KNN_m به تفکیک برای مناطق سه گانه تحت مطالعه نشان داده شده است. بیش برآورد مدل NN_m به وضوح در شکل مشخص می‌باشد این نقیصه در مدل NN_i مرتفع گشته است اما نمونه‌های با باقیمانده‌های زیاد در ظرفیت زراعی متعلق به این مدل می‌باشد. مدل‌های KNN_i و KNN_m تقریباً پراکنش یکسانی در ظرفیت زراعی دارند و در کل بهتر از مدل‌های NN می باشند. در پژمردگی دائم مدل NN_m نسبت به سایر مدل‌ها دامنه بیش‌تری از باقیمانده‌ها را نمایش می‌دهد که این امر در نمونه خاک‌های بجنورد از بقیه مناطق مشهودتر می‌باشد. و مابقی مدل‌ها دارای باقی مانده‌های تقریباً یکسان و قابل قبول می‌باشند. در نقطه ظرفیت زراعی مدل NN_i مشکل بیش برآورد مدل NN_m را ندارد و در نقطه پژمردگی دائم مدل NN_i باقیمانده‌های کمتری نسبت به مدل NN_m دارد. اما در مدل‌های KNN این درجه از تفاوت و بهبود وجود ندارد. این نتایج دلیل دیگری بر این امر است که شبکه‌های عصبی نسبت مدل‌های KNN نیاز بیش‌تری به وجود داده‌های مشابه برای اشتقاق دارند.

نتیجه گیری

تحقیق حاضر به منظور توسعه و مقایسه دو روش KNN و NNeMultistep جهت برآورد میزان رطوبت در نقاط ظرفیت زراعی و پژمردگی دائم با استفاده از نمونه‌های خاک مناطقی از شمال و شمال شرق ایران انجام شد. دیگر هدف فرعی این پژوهش تحلیل نحوه تاثیر نوع و تعداد عوامل ورودی بر عملکرد روش‌های مورد استفاده بود. برای انجام این پژوهش ۱۲۲ نمونه خاک از مناطق آمل و بابل (۱۸ نمونه)، کرج (۳۲ نمونه) و بجنورد (۷۲ نمونه) انتخاب شد. نتایج حاکی از توفیق نسبی هر دو مدل KNN و NNs در برآورد رطوبت داشتند. در کل عملکرد مدل‌های KNN از مدل‌های NNs رضایت بخش‌تر بود. این اختلاف نتایج در مدل‌هایی که صرفاً از اطلاعات اندازه ذرات خاک به عنوان متغیرهای مستقل استفاده می کردند شدیدتر بود. این امر نشان می‌دهد که مدل‌های KNN خصوصاً در شرایطی که بخواهیم با حداقل اطلاعات اندازه‌گیری شده میزان رطوبت در ظرفیت زراعی و پژمردگی دائم مدل‌سازی کنیم گزینه بسیار مناسبی می‌باشند. با افزودن چند عامل هیدرولیکی (مثلاً هدایت هیدرولیکی اشباع) به ورودی‌های مدل‌های NNs، نتایج این مدل‌ها بهبود قابل ملاحظه‌ای یافت. لذا می‌توان گفت وجود عوامل هیدرولیکی به عنوان ورودی در مدل‌های شبکه عصبی مصنوعی موجب می‌شود که آموزش این مدل‌ها با موفقیت بیش‌تری همراه شود. لازم به ذکر است که این یافته صرفاً از لحاظ تئوریک ارزشمند می‌باشد که می‌تواند به نوبه خود زمینه‌ساز تحقیقات بیش‌تری در آینده گردد. بدیهی است که در شرایط واقعی استفاده از مدل‌هایی که از خصوصیات هیدرولیکی به عنوان ورودی استفاده می‌کنند تنها هنگامی توصیه می‌شود که با دقتی قابل قبول و هزینه‌ای کم بتوان تخمین اولیه‌ای از این عوامل به عمل آورد. در مقابل افزودن عوامل چگالی ظاهری و ماده آلی تاثیر چندانی بر نتایج مدل‌های KNN

نداشت. هنگامی که از نمونه خاک‌های موجود برای اجرای هر دو روش به جای نمونه خاک‌های بانک داده هر کدام استفاده شد، نتایج مدل‌های NNs بهبود قابل ملاحظه‌ای یافت. این امر در مدل‌های KNN ناچیزتر بود. بنابراین می‌توان گفت مدل‌های KNN نسبت به نوع داده اشتقاق بسیار انعطاف پذیرتر می‌باشند و با وجود اینکه نمونه‌های آموزش بانک داده مدل‌های NNs همگی از روش جریان خروجی چند مرحله‌ای استفاده کرده بودند مدل‌های NNs به همسانی داده‌های آزمون و آموزش بسیار حساس می‌باشند. با این وجود می‌توان چنین عنوان نمود که در هنگامی که داده محلی کافی برای آموزش مدل‌های شبکه عصبی وجود ندارد استفاده از بانک داده‌ای که با استفاده از روش یک روش مشابه توسعه یافته باشد می‌تواند به عنوان گزینه‌ای مطلوب مد نظر قرار گیرد.

منابع

۱- خوشنود یزدی ع.ا. ۱۳۷۰. برآورد منحنی رطوبتی خاک از روی خصوصیات فیزیکی در برخی از خاکهای ایران. پایان نامه کارشناسی ارشد. دانشکده کشاورزی دانشگاه تهران. ۱۴۴ صفحه.

۲- عربی ز. ۱۳۸۳. پیش بینی منحنی رطوبتی با استفاده از توزیع ذرات خاک. دانشگاه آزاد واحد علوم و تحقیقات. پایان نامه کارشناسی ارشد. دانشکده کشاورزی. ۱۲۰ صفحه.

3- Bouma J. 1989. Using soil survey data for quantitative land evaluation. Adv. Soil Science. 9:177–213.

4- Lall U., and Sharma A. 1996. A nearest-neighbor bootstrap for resampling hydrologic time series. Water Resources Research. 32:679–693.

5- Lefebvre C. 2001. NeuroSolutions online help. Download the latest release from: http://www.neurosolutions.com/ downloads/ documentation .html.

6- McBratney A.B., Minasny B., Cattle S.R., and Vervoort R.W. 2002. From pedotransfer function to soil inference system. Geoderma. 109:41–73.

7- Merdun H., Cinar O., Meral R., and Apan M. 2006. Comparison of artificial neural network and regression pedotransfer functions for prediction of soil water retention and saturated hydraulic conductivity. Soil & Tillage Research. 90: 108–116.

8- Minasny B., McBratney A.B., and Bristow K.L. 1999. Comparison of different approaches to the development of pedotransfer functions for water retention curves. Geoderma. 93: 225–253.

9- Minasny B., Hopmans J.W., Harter T., Eching S.O., Tuli A., and Denton M. A. 2004. Neural networks prediction of soil hydraulic functions for alluvial soils using multistep outflow data. Sci. Soc. Am. J. 68:417–429.

10- Nemes A., Rawls W.J., and Pachepsky Y.A. 2006. Use of the nonparametric nearest Neighbor approach to estimate Soil hydraulic properties. Soil Sci. Soc. Am. J. 70:327–336.

11- Nemes A., Schaap M.G., and Wo¨sten J.H.M. 2003. Functional Evaluation of Pedotransfer Functions Derived from Different Scales of Data Collection, Soil Sci. Soc. Am. J. 67: 1093–1102.

12- Pachepsky, Y.A., Timlin, D.J., and Varallyay, G. 1996. Artificial neural networks to estimate soil water retention from easily measurable data. Soil Sci. Soc. Am. J. 60: 727–773.

13- Rawls W.J., Gish T.J., and Brakensiek D.L. 1991. Estimating soil water retention from soil physical properties and characteristics. Adv. Soil Science. 9: 213–234.

14- Schaap M.G., and Bouten W. 1996. Modeling water retention curves of sandy soils using neural networks. Water Resour. Res. 32: 3033–3040.

15- Schaap M.G., and Leij F.L. 1998. Database-related accuracy and uncertainty of pedotransfer functions. Soil Science. 10:765–779.

16- Schaap M.G., Leij F.J., and Van Genuchten M.Th. 1998. Neural network analysis for hierarchical prediction of soil hydraulic properties. Soil Sci. Soc. Am. J. 62:847–855.

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- Schumacher B.A. 2002. Methods for the determination of total organic carbon (TOC) in soils and sediments, Ecological Risk Assessment Support Center Office of Research and Development US. Environmental Protection Agency. p: 23.

19- Soil Survey Staff. 1997. National characterization data. Soil Survey Laboratory. National Soil Survey Center. and Natural Resources Conservation Service. Lincoln, NE.

20- U.S.D.A. 1982. Procedures for collecting soil samples and methods of analysis for soil survey. Soil Survey Investigations. Report No.1.

21- Wo¨sten J.H.M., Finke P.A., and Jansen M.J.W. 1995. Comparison of class and continuous pedotransfer functions to generate soil hydraulic characteristics. Geoderma. 66: 227–237.

22- Yakowitz S. 1993. Nearest-neighbor estimation for null-recurrent Markov time series. Stoch. Proc. Appl. 48:311–318.

Journal of Water and Soil
Vol. 24, No. 4, Sep-Oct 2010, p. 804-814

نشریه آب و خاک
جلد ۲۴، شماره ۴، مهر - آبان ۱۳۸۹، ص. ۸۱۴-۸۰۴

Estimating of Water Content in FC and PWP in North and North East of Iran's Soil Samples Using k-Nearest Neighbor and Artificial Neural Networks

A.Haghverdi[1]*-B.Ghahraman[2]-A.A.Khoshnood Yazdi[3]-Z.Arabi[4]

Abstract

Field capacity and permanent wilting point are the most important potential points in modeling and management of agiculutral products water requirement. Direct methods of identifying water content are time consuming and expensive. So use of pedotransfer functions for converting easily measurable characterictics to hydraulic charachteristics is a suitable solution to solve this problem. In this research efficency of NeuroMultistep outflow and K nearest neighborhood Models in deriving PTFs for identifying water content in FC and PWP for 122 soil samples of north and north east of Iran was investigated. In addition effect of different input parameters and data for deriving in both methods was specified. Results showed that in general KNN method (RMSE= 0.027) had better results than NNs models (RMSE= 0.037). Also we can say that sensitivity of NNs to quality and kind of deriving data is very high and heterogeneous data can decrease the efficiency of these models and increase the RMSE around 100%. Also the results showed that considering one or more hydraulic parameters as an input can improve the modeling results.

Key words: k-Nearest Neighbor, Artificial Neural Networks, Pedotransfer function, Field capacity, Permanent wilting point

1,2-Ph.D. Student and Professor of Irrigation and Drainage, Water Engineering Department, Ferdowsi University of Mashhad
(*-Corresponding Author Email: amirhaghverdii@gmail.com)
3-Lecturer of Shirvan College of Agriculture, Ferdowsi University of Mashhad
4-Former Graduate Student of Soil Scince.Azad University, Science and Research Center,Tehran