نشریه آبیاری و زهکشی ایران
شماره۱، جلد۸، فروردین - اردیبهشت ۱۳۹۳، ص. ۱۵۲-۱۴۵

Iranian Journal of Irrigation and Drainage
No. 1, Vol. 8, Mar-Apr 2014, p. 145-152

تخمین دمای خاک با استفاده از روش‌های نوین داده‌کاوی

لیدا اسدی*[۱]، ابوطالب هزارجریبی[۲]، خلیل قربانی[۳]، مهدی ذاکری‌نیا[۴]، زهرا آقاشریعتمداری[۵]

تاریخ دریافت: ۱۳۹۲/۸/۸ تاریخ پذیرش: ۱۳۹۳/۱۲/۲۱

چکیده

دمای خاک به‌عنوان یکی از پارامترهای مهم زیست‌محیطی، تنها در ایستگاه‌های هواشناسی سینوپتیک و فقط سه نوبت در روز اندازه‌گیری می‌شود در حالی که دیگر پارامترهای هواشناسی مانند دمای هوا و رطوبت نسبی در بیش‌تر ایستگاه‌های هواشناسی به‌جز ایستگاه‌های باران‌سنجی ثبت می‌شوند. در این پژوهش انجام شده سه روش رگرسیون چندمتغیره خطی، شبکه عصبی مصنوعی و نزدیک‌ترین k- همسایگی برای تخمین دمای اعماق مختلف خاک براساس دمای هوا و رطوبت نسبی و داده‌های روزانه ایستگاه هواشناسی سینوپتیک گرگان طی دوره ۱۰ ساله آماری ۱۹۹۶-۲۰۰۵ مورد ارزیابی قرار گرفتند. نتایج نشان دادکه با افزایش عمق، میزان تأثیر پارامترهای هواشناسی و دقت پیش‌بینی دمای خاک کاهش می‌یابد. بالا بودن دقت مدل‌ها در لایه‌های سطحی خاک به‌واسطه تأثیرپذیری بیش‌تر دمای خاک از عوامل جوی و تأخیر زمانی کم‌تر جهت انتقال حرارت از سطح به این لایه‌ها می‌باشد. مقایسه مقادیر مشاهده شده و مقادیر پیش‌بینی شده از خروجی مدل‌های مختلف نشان می‌دهد که بهترین عملکرد از اجرای الگوریتم نزدیک‌ترین k- همسایگی به‌دست می‌آید که نسبت به سایر مدل‌ها ضریب همبستگی بالاتر و میزان خطای کم‌تری دارد. پس از آن مدل‌های شبکه عصبی و رگرسیون به‌ترتیب رتبه‌های دوم و سوم را خواهند داشت.

واژه‌های کلیدی: دمای خاک، متغیرهای هواشناسی، نزدیک‌ترین k- همسایگی، شبکه عصبی مصنوعی، رگرسیون

مقدمه

دمای خاک و چگونگی تغییرات آن نسبت به زمان و مکان یکی از مهم‌ترین عواملی است که نه‌تنها تبادل ماده و انرژی را در خاک تحت تأثیر قرار می‌دهد، بلکه می‌توان گفت میزان و جهت کلیه فرآیندهای فیزیکی خاک به‌صورت مستقیم یا غیرمستقیم وابسته به دمای خاک است. دمای خاک در فرآیندهای تبخیر و تعرق، تهویه خاک، جوانه‌زنی بذر، رشد گیاه، توسعه ریشه‌ها و فعالیت میکروارگانیسم‌های خاک نقش به‌سزائی دارد. همچنین دمای خاک ارتباط تنگاتنگی با عمق نفوذ یخبندان دارد که این امر در کشاورزی، سازه‌های زهکشی و شبکه‌های آب‌رسانی شهری نقش به‌سزائی دارد. رشد بهینه گیاه بستگی به دمای مطلوب خاک دارد، به‌گونه‌ای که مقدار دمای خاک، زمان کاشت بذر را تحت تأثیر قرار می‌دهد. دمای خاک متأثر از عوامل متعددی از جمله توپوگرافی، تابش خورشید، دمای هوا، توزیع بارش، میزان رطوبت خاک، نوع و ویژگی گرمایی خاک، همانند ظرفیت گرمایی، ضریب رسانایی گرمایی و گرمای ویژه می‌باشد (نجفی‌مود و همکاران، ۱۳۸۷). سومرس و همکاران نشان دادند که گونه‌های مختلف گیاهی به دماهای متفاوتی نیازمندند و باردهی آن‌ها در محدوده معینی از درجه حرارت به بیش‌ترین مقدار می‌رسد (Sommers et al, 1981). اهمیت دمای خاک در علوم کشاورزی و هیدرولوژی ازیک سو و وجود دشواری‌های فراوان در ثبت این پارامتر حیاتی پژوهش‌گران را برآن داشته است تا به‌دنبال یافتن رابطه‌ای بین دمای خاک و پارامترهای دیگر باشند تا بتوانند دمای خاک را با دقت مطلوب برآورد کنند. تاکنون مطالعات زیادی در زمینه برآورد دمای خاک صورت گرفته است. ماکلین و آیرتس با استفاده از رگرسیون خطی چندمتغیره ارتباط بین دمای خاک با پارامترهای هواشناسی را تعیین و با ارائه مدل به تخمین دمای خاک در اعماق ۱، ۲، ۵، ۱۰ سانتی‌متری پرداختند. متغیرهای مورد استفاده شامل دمای هوا و درصد ابرناکی در روز، دمای هوا در دو روز قبل، نسبت ساعات آفتابی و روز فصل بود (Maclean & Ayers,1985). آلدریچ و کوک با

۱- دانشجوی کارشناسی ارشد آبیاری و زهکشی، گروه مهندسی آب، دانشگاه علوم کشاورزی و منابع طبیعی
۲- دانشیار گروه مهندسی آب، دانشگاه علوم کشاورزی و منابع طبیعی
۳ و ۴- استادیار گروه مهندسی آب، دانشگاه علوم کشاورزی و منابع طبیعی
۵- استادیار گروه مهندسی آب، دانشگاه تهران
(* - نویسنده مسئول: (Email: lidaasadi69@yahoo.com

ارائه مدل رگرسیون چندمتغیره، میانگین سالیانه دمای خاک در عمق های ۱۰ و ۳۰ سانتی‌متر را محاسبه کردند (Aldridge & Cook,1983). میهالاکاکو در آتن و دوبلین برای برآورد تغییرات دمای خاک از تاب تشخیص و شبکه عصبی مصنوعی با دمای هوا، رطوبت نسبی و تابش خورشیدی به‌عنوان متغیرهای ورودی استفاده نمود (Mihalakakou,2002). مایکل و گیلچریست با ارائه مدل چندجمله‌ای درجه چهارم موفق به محاسبه کمینه و بیشینه دمای خاک در هر روز از سال در ۱۸ منطقه از انگلستان شدند(Meikle & Gilchrist, 1983). با توجه به اهمیت دسترسی به داده‌های دمای خاک، در کشور نیز پژوهش‌هایی در خصوص برآورد دمای خاک صورت گرفته است. ثنایی و همکاران (۱۳۸۷) با استفاده از تحلیل سری‌های فوریه و دوره آماری ۱۲ ساله به برآورد دمای خاک روزانه و سالیانه در عمق‌های مختلف در ۶ ایستگاه کردستان اقدام نمودند. در این تحلیل آن‌ها تنها با استفاده از دمای هوا (۲ متری) و بدون در نظر گرفتن دیگر پارامترهای هواشناسی (تابش، ساعات آفتابی، رطوبت هوا، ابرناکی) روابط همبستگی خود را جهت برآورد دمای خاک پیشنهاد نمودند. ابراهیمی (۱۳۷۴) با استفاده از روش‌های آماری به مطالعه دمای اعماق مختلف خاک در تهران و همدان پرداخت. جعفری‌گلستانی و همکاران (۱۳۸۶) نیز با استفاده از روابط رگرسیونی، معادلات تجربی را جهت تخمین دمای خاک شهر ساری در عمق‌های ۷ و ۲۰ سانتی‌متری پیشنهاد نمودند. در تحقیق دیگری، نجفی‌مود و همکاران (۱۳۸۷) نیز به‌منظور برآورد عمق یخبندان خاک و ارائه یک رابطه ساده بین درجه حرارت هوا و دمای خاک و اعماق مختلف آن، مطالعه‌ای را در شهرستان‌های استان خراسان رضوی انجام دادند. مهدویان (۱۳۷۸) جهت پیش‌بینی دمای خاک در منطقه کرج از شبکه عصبی مصنوعی و روابط تجربی برانت استفاده نمود. سبزی‌پرور و همکاران (۱۳۸۹) با مطالعه روی داده‌های هواشناسی ۳ ایستگاه سینوپتیک زاهدان، تهران و رامسر که دارای اقلیم متفاوتی می‌باشند، نتیجه گرفتند که دقت روش استنتاج تطبیقی عصبی–فازی ۴ درصد بیش‌تر از روش آماری رگرسیون در پیش‌بینی دمای خاک در اعماق مختلف بوده است. همچنین دقت پیش‌بینی دمای خاک با روش استنتاج تطبیقی عصبی–فازی در دو ایستگاه زاهدان و تهران (اقلیم خشک) به‌ترتیب ۱۲ درصد و ۴/۵ درصد بهتر از رامسر (اقلیم مرطوب) مشاهده گردید.

همان‌طوری که در سوابق تحقیق مشاهده شده است از دیرباز تخمین دمای خاک مورد توجه محققان بوده و در این خصوص با روش‌هایی مثل سری فوریه، موازنه انرژی و ... مورد استفاده قرار گرفته و به‌ندرت روش‌های جدیدی چون k– همسایگی، شبکه عصبی و نیز مقایسه این دو روش در برآورد دمای خاک مورد توجه قرار گرفته است. بنابراین در تحقیق حاضر خلأ پژوهشی مذکور مدنظر قرار گرفته و با استفاده از آمار و اطلاعات موجود در ایستگاه هواشناسی سینوپتیک گرگان و همچنین الگوریتم‌های مدل‌سازی داده‌کاوی شامل روش شبکه عصبی مصنوعی و نزدیک‌ترین k– همسایگی و روش رگرسیون چندمتغیره، روابط موجود بین داده‌های دمای هوا و رطوبت نسبی با دمای خاک در اعماق مختلف جستجو شده و بهترین روش معرفی می‌گردد.

مواد و روش‌ها

منطقه مطالعاتی این تحقیق ایستگاه هواشناسی سینوپتیک گرگان می‌باشد که دارای موقعیت جغرافیایی به طول ۵۴ درجه و ۲۴ دقیقه، عرض جغرافیایی ۳۶ درجه و ۵۴ دقیقه می‌باشد. شهر گرگان با وسعت حدود ۶۰ کیلومتر مربع و در شمال ایران با آب و هوای معتدل و مرطوب قرار دارد. به‌منظور بررسی دمای خاک، داده‌های روزانه دمای اعماق مختلف خاک (اعماق ۵، ۲۰، ۳۰، ۵۰ و ۱۰۰ سانتی‌متری) در ساعات ۰۳، ۰۹، ۱۵ در دوره آماری ۱۰ ساله (۲۰۰۵–۱۹۹۶) از سازمان هواشناسی کشور تهیه شد. دمای خاک در اعماق مختلف به‌عنوان متغیر وابسته، دمای هوا و رطوبت نسبی همراه با تأخیرهای زمانی سه و شش ساعت به‌عنوان متغیرهای مستقل به روش‌های مدل‌سازی معرفی شدند. لازم به‌ذکر است که ۸۰ درصد داده‌ها را جهت آموزش و ۲۰ درصد دیگر را برای آزمون داده‌ها در نظر گرفته شد. نتایج این مدل‌سازی به‌شرح زیر می‌باشد:

روش‌های آماری تخمین دمای خاک

در این پژوهش سه روش آماری رگرسیون چندمتغیره خطی، شبکه عصبی مصنوعی و نزدیک‌ترین k– همسایگی جهت تخمین دمای خاک مورد ارزیابی قرار گرفته‌اند.

روش رگرسیون چندمتغیره خطی

در این روش بر اساس داده‌های ورودی و خطوط تصمیم‌گیری، داده‌های خروجی پیش‌بینی می‌شوند. در واقع تکنیک رگرسون، یک مجموعه‌ای از داده‌های آزمایشی را دریافت نموده و فرمولی را به آن داده‌ها برازش می‌نماید. رگرسیون می‌تواند جهت حل مسائل کلاس‌بندی و نیز جهت پیش‌گویی مورد استفاده قرار بگیرد. روش رگرسیون خطی، یک تکنیک یادگیری نظارتی است که به‌وسیله آن می‌خواهیم تغییرات یک متغیر وابسته به‌وسیله ترکیب خطی از یک یا چند متغیر مستقل مدل کنیم. حالت کلی معادله آن به این صورت است، $f(x_1+x_2+\ldots+x_n)=a_1x_1+a_2x_2+\ldots+a_nx_n+b$ که xها متغیر مستقل و a و b ضرایب ثابت هستند و $f(x_1.x_1\ldots x_n)$ متغیر وابسته می‌باشند.

روش شبکه عصبی مصنوعی

شبکه عصبی مصنوعی (Ann[1]) با تجزیه و تحلیل داده‌های ورودی و نتایج نظیر آن‌ها ارتباطی منطقی بین داده‌ها برقرار می‌کند که ممکن است غیرخطی و نامشخص باشد، سپس با استفاده از این ارتباط منطقی کار شبیه‌سازی را برای موارد احتمالی مشابه انجام می‌دهد. این شبکه‌ها بر مبنای مقایسه خروجی شبکه و هدف، تعدیل می‌شوند و تا زمانی که خروجی شبکه با خروجی هدف تطبیق پیدا کند این کار ادامه می‌یابد. شبکه‌های آموزش‌یافته را می‌توان به‌صورت رویه‌ای از جعبه سیاه برای برآوردهای غیرخطی با عنوان نگاشت‌های غیرخطی قابل تنظیم معرفی کرد، چرا که فضای بردار ورودی را به‌وسیله مجموعه‌ای از توابع غیرخطی به فضای خروجی مرتبط می‌سازند. کاربردی‌ترین نوع شبکه‌های عصبی، شبکه پرسپترون چندلایه (MLP) و شبکه‌های توابع شعاعی (RBF) می‌باشند (Salehi et al, 1998). در این تحقیق برای شبیه‌سازی سرعت باد با استفاده از شبکه‌ی MLP، از دو لایه‌ی پنهان با تعداد نرون‌های متفاوت و توابع محرک کران‌دار تابع سیگموئید و از میان روش‌های مختلف آموزش به روش پس‌انتشار خطا با الگوریتم لورنبرگ – مارکوارت، به‌دلیل همگرایی سریع‌تر در آموزش شبکه استفاده شده است (قبائی‌سوق و همکاران، ۱۳۸۹). اساس روش پس‌انتشار خطا بر پایه قانون یادگیری اصلاح خطا می‌باشد که از دو مسیر اصلی رفت و برگشت تشکیل می‌شود. در مسیر رفت، بردار ورودی به شبکه اعمال شده و تأثیراتش از طریق لایه‌های میانی به لایه خروجی انتشار می‌یابد و بردار خروجی پاسخ واقعی شبکه را تولید می‌نماید. شبکه عصبی مصنوعی شامل مجموعه‌ای از گره‌ها (گره‌ها، نرون‌ها، عناصر محاسباتی) بوده به‌گونه‌ای که هر گره دارای ورودی و خروجی بوده و هر گره بر اساس تابعی خاص محاسبه ساده‌ای انجام داده و بین گره‌ها، اتصالات موزون بر اساس معماری شبکه مشخص وجود دارد. نتیجه یک شبکه عصبی مصنوعی تابعی بسیار پیچیده از ارتباطات موزون می‌باشد.

روش نزدیک‌ترین k- همسایگی (KNN[2])

الگوریتم نزدیکترین k- همسایگی جزء روش‌های یادگیری نظارت‌شده است که هم در طبقه‌بندی و هم در پیش‌بینی مورد استفاده قرار می‌گیرد. نحوه عملکرد این الگوریتم براساس مشاهدات و نمونه‌ها می‌باشد. بر اساس این الگوریتم می‌توان یک نمونه جدید را بر اساس اکثریت k گروه و دسته که نزدیک‌ترین همسایگی‌ها را با آن نمونه داشته باشند، تقسیم‌بندی نمود. به‌عبارت دیگر می‌توان گفت این روش، k تعداد از الگوهای مشابه را پیدا کرده و بر اساس آن‌ها ارزش نمونه مورد مطالعه را پیش‌بینی می‌کند. در روش نزدیک‌ترین k- همسایگی نمونه ناشناخته S در یک کلاس از پیش تعریف شده C_i متعلق به C بر اساس نمونه‌ها کلاس‌بندی شده قبلی (داده‌های آموزش) طبقه‌بندی می‌شود. زمانی که نمونه S کلاس‌بندی می‌شود، KNN فاصله آن را با همه نمونه‌های موجود در داده‌های آموزشی اندازه‌گیری می‌کند. فاصله اقلیدسی از متداول‌ترین معیارهای اندازه‌گیری فاصله می‌باشد. سپس تمام مقادیر فاصله مرتب می‌شوند به‌طوری که $d_i \le d_{i+1}$; i = 1, 2, 3, …, n می‌باشد. k نمونه با کم‌ترین فاصله به نمونه جدید به‌عنوان نزدیک‌ترین k- همسایه شناخته می‌شوند و برای کلاس‌بندی نمونه جدید S به کلاس موجود $C_i \in C$, $1 < i \le n$ استفاده می‌شود. تصمیم طبقه‌بندی به طبیعت داده‌ها بستگی دارد. تصمیم طبقه‌بندی به طبیعت داده‌ها بستگی دارد. این الگوریتم جزء روش‌های تنبل به‌حساب می‌آید به این دلیل که مرحله آموزش را همان زمان که نمونه جدید می‌بایست طبقه‌بندی شود اجراء می‌کند برخلاف الگوریتم‌های یادگیری، نقطه مقابل آن که داده‌های آموزشی را قبل از ورود نمونه جدید طبقه‌بندی می‌کنند. بر این اساس الگوریتم KNN نسبت به دیگر الگوریتم‌های یادگیری محاسبات بیش‌تری را لازم دارد. KNN برای داده‌های پویا و داده‌هایی که سریع تغییر کرده و به‌روز می‌شوند، مناسب است (جان و همکاران، ۲۰۰۸).

ارزیابی و مقایسه مدل‌های پیش‌بینی

با در نظر گرفتن مقادیر مشاهده شده X و مقادیر پیش‌بینی شده Y توسط هر یک از مدل‌ها، از دو معیار ریشه میانگین مربعات خطا (RMSE) و میانگین خطای اریبی (MBE) برای ارزیابی مدل‌ها استفاده شد. معیار (RMSE) بزرگی خطا و معیار (MBE) میزان انحراف از خط نیمساز را نشان می‌دهد. همچنین ضریب تعیین نیز برای نشان دادن میزان همبستگی بین مقادیر مشاهده شده و پیش‌بینی شده محاسبه شد.

$$R^2 = \frac{\left[\sum_{i=1}^{n}(X_i-\bar{X})(Y_i-\bar{Y})\right]^2}{\left[\sum_{i=1}^{n}(X_i-\bar{X})^2\right]\left[\sum_{i=1}^{n}(Y_i-\bar{Y})^2\right]} \quad (۱)$$

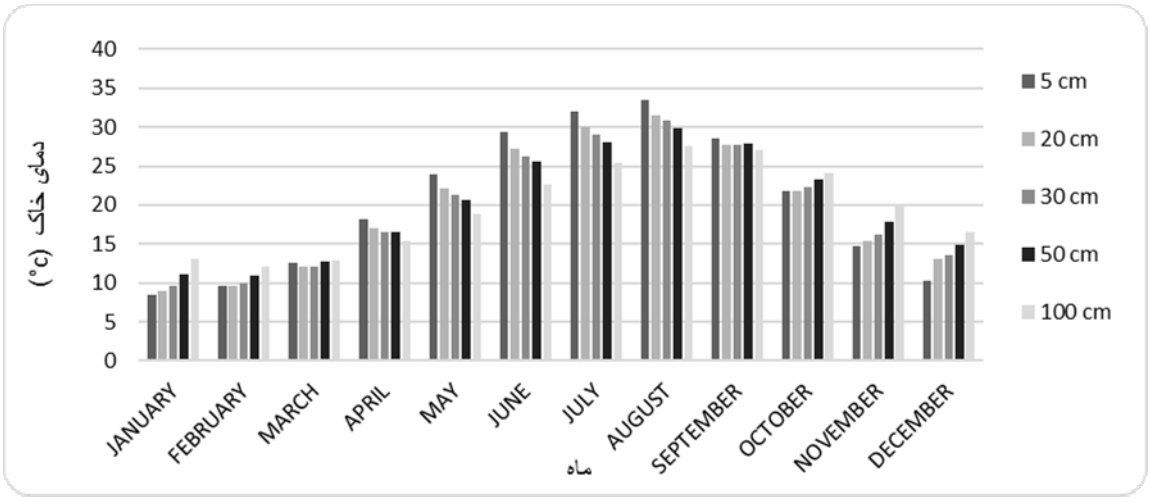
$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(X_i-Y_i)^2}{n}} \quad (۲)$$

$$MBE = \frac{\sum_{i=1}^{n}(X_i-Y_i)}{n} \quad (۳)$$

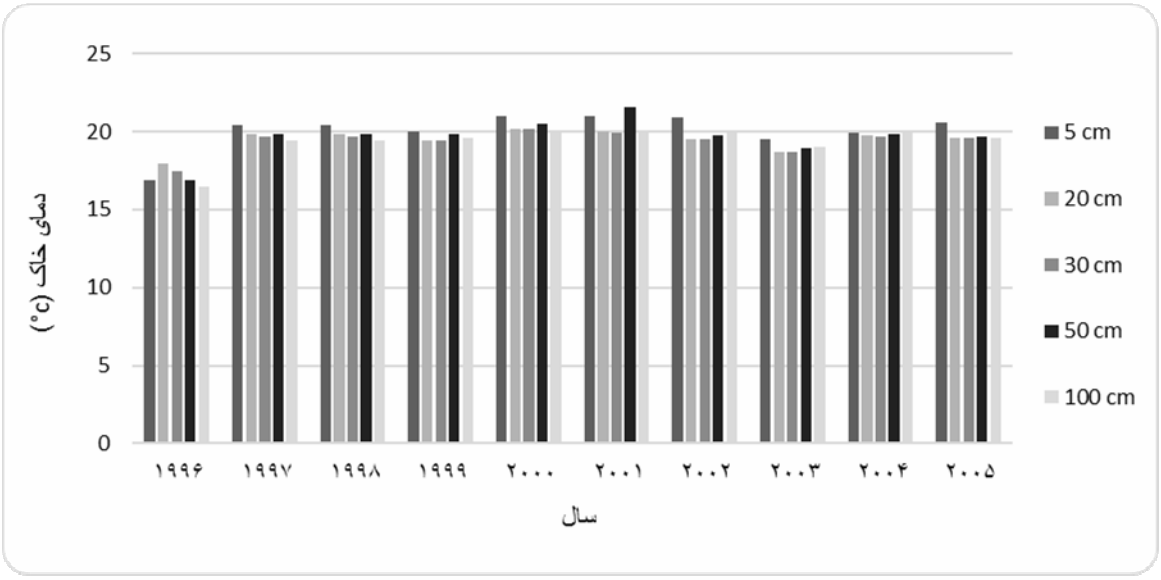
تغیرات ماهانه و سالانه دمای خاک

شکل (۱) تغییرات میانگین ماهانه دمای واقعی خاک را در کل دوره آماری در شهر مورد نظر نشان می‌دهد. همان‌طور که ملاحظه می‌گردد، حداکثر دمای ماهانه خاک در همه اعماق در ماه آگوست (مرداد) اتفاق افتاده است. حداقل مقدار آن نیز در اعماق کم (۵ و ۱۰ سانتی‌متری) در ماه ژانویه (دی) و در بقیه اعماق در ماه فوریه (بهمن) حادث گردیده است.

1- Artificial Neural Network
2- K-Nearest Neighbor



شکل۱- تغییرات ماهانه دمای خاک در دوره آماری ۲۰۰۵-۱۹۹۶



شکل۲- تغییرات سالانه دمای خاک در دوره آماری ۲۰۰۵-۱۹۹۶

در شکل (۲)، تغییرات سالانه دمای خاک در کل دوره آماری نشان داده شده است. همان‌طور که انتظار می‌رود، مقادیر میانگین سالانه دمای خاک در طول دوره مورد مطالعه دارای نوساناتی است که این امر ناشی از تأثیر دیگر متغیرهای جوی مانند: تابش خورشیدی، ابرناکی، ساعات آفتابی، رطوبت هوا، بارش و ویژگی‌های فیزیکی خاک بر دمای خاک می‌باشد.

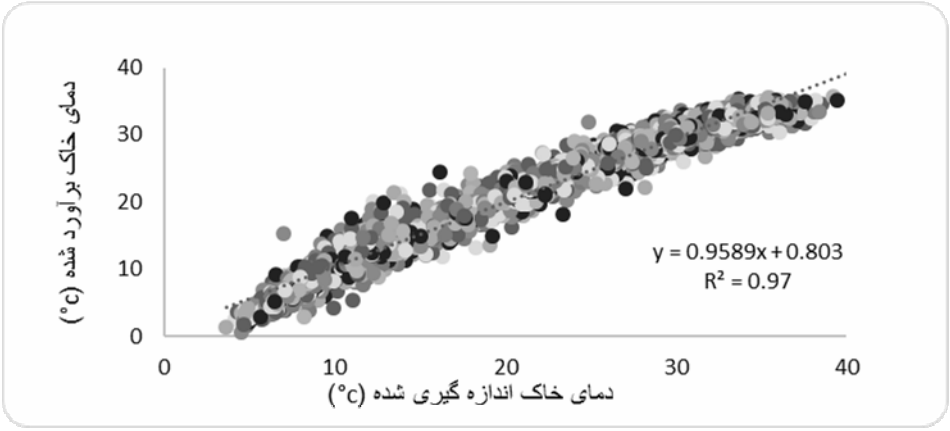
نتایج و بحث

ارزیابی مدل‌های پیش‌بینی در تخمین دمای خاک

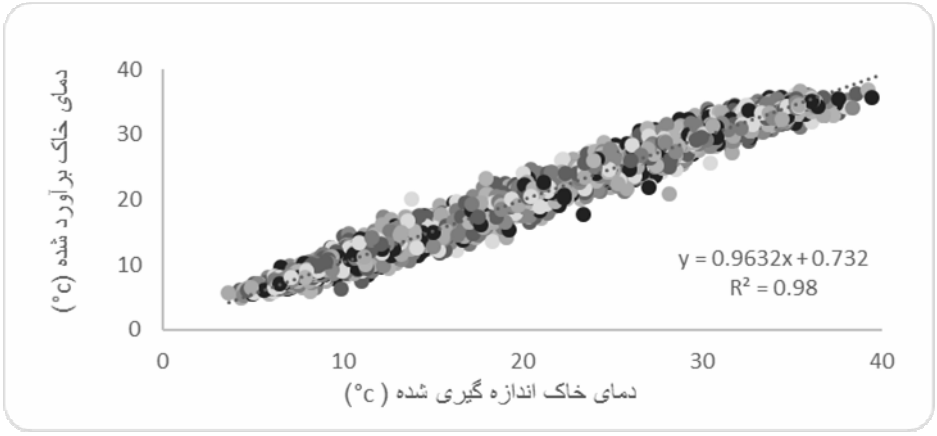
با توجه به داده‌های موجود در مورد دمای خاک و مقایسه آن‌ها با مقادیر برآورد شده توسط مدل‌ها، نتیجه گرفته شد که مدل نزدیک‌ترین k- همسایگی نسبت به دیگر مدل‌های مورد استفاده در این پژوهش، دمای اعماق مختلف خاک را با دقت بالاتری برآورد می‌کند. با این توضیحات با توجه به شکل‌های ۳، ۴ و ۵ که مقادیر اندازه‌گیری شده و برآوردشده توسط مدل‌ها را نشان می‌دهد، می‌توان مشاهده کرد که در مدل نزدیک‌ترین k- همسایگی داده‌ها پراکندگی مناسب‌تری در اطراف خط رگرسیونی دارند. برای رسیدن به مدل مناسب از معادله خطی رگرسیون ($y = ax + b$) استفاده می‌گردد. هر چه ضریب متغیر x به یک نزدیک‌تر و مقدار ثابت b به صفر نزدیک شود زاویه خط برازش داده با محور افق به ۴۵ درجه نزدیک‌تر می‌گردد. البته اشکال ۳، ۴ و ۵ فقط برای عمق ۵ سانتی‌متری آورده

شده است اما توضیحات فوق برای کلیه‌ی اعماق صدق می‌کند.

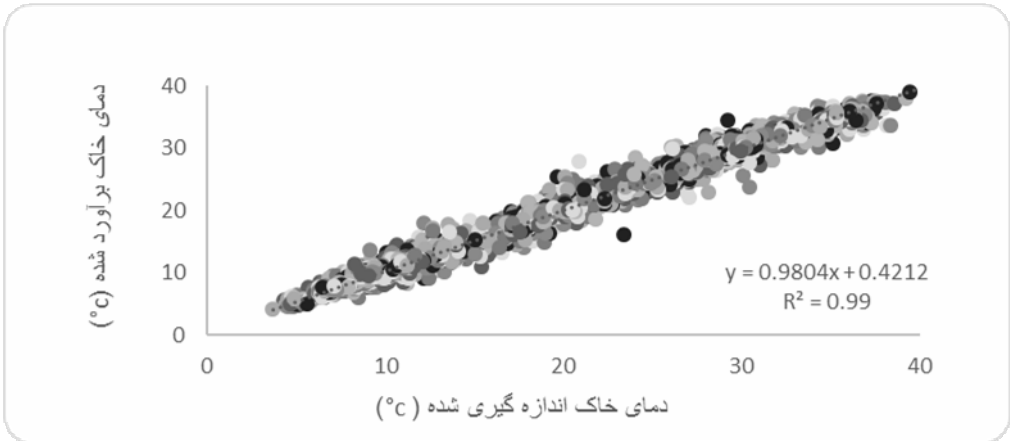
نتایج خروجی مدل‌های پیش‌بینی نشان می‌دهد که با افزایش عمق خاک دقت برآورد دمای خاک بر اساس داده‌های هواشناسی کاهش می‌یابد و این می‌تواند ناشی از عوامل تأثیر شرایط خاک مانند وضعیت رطوبتی خاک بر انتقال حرارت در خاک باشد که متأسفانه در ایستگاه‌های هواشناسی اندازه‌گیری نمی‌شود.



شکل ۳- مقادیر مشاهده شده و پیش‌بینی شده دمای عمق ۵ سانتی‌متری خاک با استفاده از مدل رگرسیون چندمتغیره خطی



شکل ٤- مقادیر مشاهده شده و پیش‌بینی شده دمای عمق ۵ سانتی‌متری خاک با استفاده از مدل شبکه عصبی مصنوعی



شکل ۵- مقادیر مشاهده شده و پیش‌بینی شده دمای عمق ۵ سانتی‌متری خاک با استفاده از مدل نزدیک‌ترین k- همسایگی

جدول ۱- معیارهای ارزیابی خطا و مقادیر R^2 برای داده‌های آموزش و آزمون مدل رگرسیون چندمتغیره خطی

اعماق خاک	۵		۲۰		۳۰		۵۰		۱۰۰	
	آموزش	آزمون	آموزش	آزمون	آموزش	آزمون	آموزش	آزمون	آموزش	آزمون
RMSE	۱/۹۹	۲/۰۱	۱/۹۰	۲/۰۴	۲/۰۸	۲/۰۸	۲/۳۹	۲/۴۰	۲/۶۵	۲/۶۹
MBE	-۰/۰۲۶	-۰/۰۵۷	-۰/۰۲۴	۰/۰۲۰	-۰/۰۲۳	۰/۰۲۰	-۰/۰۲۷	۰/۰۵۳	۰/۰۰۲	-۰/۰۱۷
R^2	۰/۹۷۶	۰/۹۷۶	۰/۹۷۱	۰/۹۷۱	۰/۹۶۳	۰/۹۶۰	۰/۹۳۸	۰/۹۳۶	۰/۸۸۰	۰/۸۷۵

جدول ۲- معیارهای ارزیابی خطا و مقادیر R^2 برای داده‌های آموزش و آزمون مدل شبکه عصبی مصنوعی

اعماق خاک	۵		۲۰		۳۰		۵۰		۱۰۰	
	آموزش	آزمون	آموزش	آزمون	آموزش	آزمون	آموزش	آزمون	آموزش	آزمون
RMSE	۱/۶۹	۱/۷۷	۱/۶۱	۱/۶۶	۱/۷۶	۱/۸۶	۲/۱۰	۲/۲۰	۲/۴۲	۲/۵۳
MBE	-۰/۰۰۷	-۰/۰۵۰	۰/۰۱۳	۰/۰۴۱	۰/۰۹۰	۰/۰۱۱۰	۰/۰۶۶	۰/۱۲۹	۰/۱۷۵	۰/۱۲۹
R^2	۰/۹۸۳	۰/۹۸۱	۰/۹۷۹	۰/۹۷۸	۰/۹۷۳	۰/۹۶۹	۰/۹۵۳	۰/۹۴۷	۰/۹۰۲	۰/۸۹۱

جدول ۳- معیارهای ارزیابی خطا و مقادیر R^2 برای داده‌های آموزش و آزمون مدل نزدیکترین k- همسایگی

اعماق خاک	۵		۲۰		۳۰		۵۰		۱۰۰	
	آموزش	آزمون	آموزش	آزمون	آموزش	آزمون	آموزش	آزمون	آموزش	آزمون
RMSE	۱/۲۹	۲/۱۷	۱/۲۲	۲/۰۴	۱/۳۲	۲/۲۰	۱/۵۷	۲/۶۵	۱/۷۴	۳/۰۵
MBE	۰/۰۱۴	۰/۰۵۲	۰/۰۱۹	۰/۰۸۴	-۰/۰۰۲	۰/۰۴۳	-۰/۰۱۳	۰/۰۵۱	۰/۰۰۹	-۰/۰۵۶
R^2	۰/۹۹۱	۰/۹۷۲	۰/۹۸۹	۰/۹۶۸	۰/۹۸۵	۰/۹۵۷	۰/۹۷۴	۰/۹۲۴	۰/۹۵۱	۰/۸۴۳

جداول (۱)، (۲) و (۳) آنالیز حساسیت و مقادیر ضریب رگرسیونی برای داده‌های آموزش و آزمون مدل‌ها را نشان می‌دهد. همان‌طور که مشاهده شد در این جداول مقادیر بزرگی خطا برای مدل نزدیک‌ترین k- همسایگی کم‌تر از سایر مدل‌ها بوده و نشان‌دهنده بالاتر بودن دقت این مدل نسبت به بقیه مدل‌ها است.

توجه به مقادیر منفی خطای MBE به‌دست آمده در جداول در اغلب عمق‌ها، حاکی از فروبرآورد (Underestimation) دمای خاک توسط مدل رگرسیونی به شکل نامحسوسی دارد.

نتیجه‌گیری

در پژوهش حاضر، دمای خاک ایستگاه سینوپتیک شهر گرگان با آب و هوای معتدل با استفاده از سه الگوریتم رگرسیون‌گیری، نزدیک‌ترین k- همسایگی و شبکه عصبی مصنوعی با در نظر گرفتن متغیرهای هواشناسی مورد بررسی قرار گرفت. بررسی آماره‌های واسنجی نشان داد که با افزایش عمق، میزان تأثیر متغیرهای هواشناسی بر دمای خاک کاهش می‌یابد. بالا بودن دقت مدل‌ها در لایه‌های سطحی خاک به‌واسطه تأثیرپذیری بیش‌تر دمای خاک از عوامل جوی و تأخیر زمانی کم‌تر جهت انتقال حرارت از سطح به این لایه‌ها می‌باشد. نمودارهای ارائه شده در اعماق سطحی (۵ و ۲۰ سانتی‌متری) بیش‌ترین همبستگی و در ۱۰۰ سانتی‌متری کم‌ترین همبستگی را با متغیرهای هواشناسی ارائه نمودند. با تطابق مقادیر مشاهده شده نسبت به مقادیر پیش‌بینی شده از خروجی نرم‌افزار نمودار نشان می‌دهد که بهترین عملکرد از اجرای الگوریتم نزدیک‌ترین k- همسایگی به‌دست می‌آید که نسبت به سایر مدل‌ها ضریب همبستگی بالاتر و میزان خطای کم‌تری دارد. پس از آن مدل‌های شبکه عصبی و رگرسیون به‌ترتیب رتبه‌های دوم و سوم را خواهند داشت. در مقایسه این پژوهش با نتایج ارائه شده توسط دیگر پژوهشگران می‌توان نتیجه گرفت که در مورد شبکه عصبی و رگرسیون نتایج این پژوهش با نتایج تحقیق جعفری‌گلستانی و همکاران (۱۳۸۶) که استفاده از روابط رگرسیونی، معادلات تجربی را جهت تخمین دمای خاک شهر ساری در عمق‌های ۷ و ۲۰ سانتی‌متری پیشنهاد نمودند و همچنین مهدویان (۱۳۷۸) که جهت پیش‌بینی دمای خاک در منطقه کرج از شبکه عصبی مصنوعی و روابط تجربی برانت استفاده نمود مطابقت دارد ولی در روش نزدیکترین k- همسایگی به‌عنوان یکی از روش‌های نوین در داده‌کاوی که وجه تمایز این مطالعه می‌باشد گزارشی یافت نشده است تا مقایسه‌ای صورت گیرد.

منابع

ابراهیمی،ا. ۱۳۷۴. مطالعه دمای اعماق مختلف خاک. پایان‌نامه کارشناسی ارشد هواشناسی، مؤسسه ژئوفیزیک دانشگاه تهران.

ثنائی نژاد،ح.، ادیب‌عباسی،م.، موسوی‌بایگی،م و حیدری‌گندمان،م.ط. ۱۳۸۷. بررسی رژیم دمایی هوا و اعماق خاک و تعیین توابع نوسانات ادواری آن‌ها در ایستگاه های کردستان. مجله علوم و صنایع کشاورزی، ویژه آب و خاک ۲۲:۲۵-۳۳.

جعفری‌گلستانی،م.، رائینی‌سرجاز،م و ضیاءتباراحمدی،م. ۱۳۸۶. برآورد دمای ژرفای خاک با بهره‌گیری از روش تجزیه منحنی و همبستگی‌های رگرسیونی برای شهرستان ساری. مجله علوم کشاورزی و منابع طبیعی (ویژه‌نامه زراعت و اصلاح نباتات)، ۱(۵): ۱۱۲ تا ۱۲۳.

سبزی پرور،ع.ا.، زارع‌ابیانه،ح و بیات‌ورکشی،م. ۱۳۸۹. مقایسه یافته-های مدل شبکه استنتاجی عصبی-فازی با مدل‌های رگرسیونی به‌منظور برآورد دمای خاک در سه اقلیم متفاوت نشریه آب و خاک ۱۴(۵۲):۱۲۵-۱۳۷.

قبائی‌سوق،م.، مساعدی،ا.، حسام،م و هزارجریبی،ا. ۱۳۸۹. ارزیابی تاثیر پیش‌پردازش پارامترهای ورودی به شبکه عصبی مصنوعی با استفاده از روش‌های رگرسیون گام‌به‌گام و گاما تست به‌منظور تخمین سریعتر تبخیر و تعرق روزانه. نشریه آب وخاک (۳):۶۲۴-۶۱۰.

مهدویان،م. ۱۳۷۸. پیش‌بینی محلی دمای حداقل شبانه‌روزی در منطقه کرج. پایان‌نامه کارشناسی ارشد، دانشگاه تهران.

نجفی‌مود،م.ح.، علیزاده،ا.، محمدیان،ا و موسوی،ج. ۱۳۸۷. بررسی رابطه دمای هوا و دمای اعماق مختلف خاک و برآورد عمق یخبندان (مطالعه موردی استان خراسان رضوی). مجله علمی-پژوهشی آب و خاک (علوم و صنایع کشاورزی)، ۲۲(۲): ۴۵۶ تا ۴۶۶.

Aldridge,R and Cook,FJ. 1983. Estimation of soil temperatures at 0.1m and 0.3m depths. New Zealand Soil Bureau Scientific Report 62: 18.

Jan,Z., AbrarM., Bashir,Sh And Mirza,A. 2008. Seasonal to inter-annual climate prediction using data mining KNN technique. In the international multi topic conference (IMTIC08) in the revised selected papers, Jamshoro, Pakistan.

Maclean,Jr.SF and Ayers, M.p. 1985. Estimation of soil temperature from climate variables at Barrow, Alaska., USA. Article and Alpine Research. 17,P. 425-432.

Meikle,R.W and Gilchrist AJ, 1983. A mathematical method for estimation of soil temperatures in England and Scotland. Agricultural Meteorology 30(3): 221-225.

Mihalakakou, G. 2002. On estimating soil surface temperature profiles. Energy and Buil. 34: 251-259.

Salehi,F., Lacroix,R and Wade,K.M. 1998. Effects of Learning Parameters and Data Presentation on the Performance of Back Propagation Networks for Milk Yield Prediction. Transaction of ASAE. 41 (1). 253-.952

Sommers,L.E., Gilmour,C.M., Wildung,R.E and Beck,S.M. 1981. The effect of water potential on decomposition processes in soils, in Water Potential Relations in Soil Microbiology. Edited by J.E. Parr, W.R. Gardner and W.R. Elliot, SSSA Spec. Publ. 9: 97-117.

Iranian Journal of Irrigation and Drainage
No. 1, Vol. 8, Mar-Apr 2014, p. 145-152

نشریه آبیاری و زهکشی ایران
شماره۱، جلد۸، فروردین - اردیبهشت ۱۳۹۳، ص. ۱۵۲-۱۴۵

Estimating Soil Temperature Using Modern Methods of Data Analysis

L.Asadi[1*], A.Hezarjaribi[2], Kh.Ghorbani[3], M.Zakernia[4], Z.AghaShariatmadari[5]
Received: Oct.30,2013 Accepted: Mar.12,2014

Abstract

Soil temperature as one of the important environmental parameters, is measured only in synoptic weather stations and three times per day, while the other meteorological parameters such as air temperature and relative humidity are measured most in meteorological stations except rain guage stations. In this research which based on three methods of multivariate linear regression, artificial neural network and k-nearest neighbor were evaluated to estimate the soil temperature at different depth based on air temperature and relative humidity using daily data from synoptic meteorology stations of Gorgan from 1996 to 2005. The results showed that with increasing soil depth, the effect of meteorological parameters and estimation accuracy will decrease.more accuracy of shallow depth soil temperature is due to the greater influence of climatic factors on soil temperature and less time delay of heat transfer from the surface to the deeper depths. By comparing the observed and predicted soil temperature values from the various models, it will be concluded that the best performance is obtained from the k-nearest neighbor algorithm which has higher R^2 and less RMSE. Artificial neural network and regression models, are in second and third place to predict soil temperature.

Keywords: Soil temperature, Meteorological parameters, K-nearest neighbor, Artificial neural network, Regression

1- MSc Student of Irrigation & Drainage Engineering, Agricultural Sciences and Natural Resources University of Gorgan
2- Associate Professor of Department of Water Engineering, Agricultural Sciences and Natural Resources University of Gorgan
3,4- Assistant Professor of Department of Water Engineering, Agricultural sciences and Natural Resources University of Gorgan
5- Assistant Professor of Department of Water Engineering, University of Tehran
(*-Corresponding Author Email: lidaasadi69@yahoo.com)