

استفاده از روش‌های نوین برای تعیین ورودی‌های موثر در تخمین دمای خاک

بابک محمدی*

۱- دانشجوی دکتری هواشناسی کشاورزی، گروه مهندسی آبیاری و آبادانی، دانشکده مهندسی و فناوری کشاورزی، دانشگاه تهران، کرج

(تاریخ دریافت: ۱۴۰۰/۰۵/۰۹، تاریخ پذیرش: ۱۴۰۱/۰۸/۳۰)

چکیده

در این تحقیق تخمین دمای عمق ۱۰ سانتی متری خاک ایستگاه سینوپتیک تبریز واقع در استان آذربایجان شرقی با استفاده از شبکه عصبی مصنوعی (ANN) و ماشین بردار پشتیبان (SVM) انجام گرفته است. از دو روش تحلیل مؤلفه اصلی (PCA) و آزمون گاما (GT) برای پیش‌پردازش داده‌ها و تعیین داده‌های ورودی استفاده گردید. با توجه به نتایج، برای ایستگاه تبریز ۳ متغیر ورودی توسط آزمون گاما انتخاب شد. در روش تحلیل مؤلفه‌های اصلی نیز برای ایستگاه سینوپتیک تبریز چهار مؤلفه اصلی انتخاب شده است. نتایج مدل‌سازی حاکی از این است که مدل ماشین بردار پشتیبان مبتنی بر آزمون گاما (GT-SVM) با مجذور میانگین مربعات خطای ۲/۴۸ درجه سلسیوس را می‌توان به عنوان مدل منتخب برای ایستگاه مذکور انتخاب کرد. دوره آماری داده‌ها از سال ۱۳۹۲ تا ۱۳۹۴ بوده است که در این دوره آماربندی تغییرات مهم شناخته شده برای تخمین دمای خاک توسط آزمون گاما به ترتیب دمای متوسط هوا، ساعت آفتابی، سرعت باد و رطوبت نسبی بودند. با توجه به نتایج می‌توان چنین نتیجه گرفت که از نظر شاخص‌های آماری به کار گرفته شده، روش‌های استفاده شده پیش-پردازش داده‌ها در این تحقیق برای پیش‌بینی دمای خاک با یکدیگر اختلاف کمی دارند و هر دو روش به خوبی عمل کرده‌اند. همچنین مدل ماشین بردار پشتیبان در تمام تخمین‌ها نسبت به مدل شبکه عصبی مصنوعی عملکرد قابل قبول تری داشته است.

کلمات کلیدی: استان آذربایجان شرقی، آزمون گاما، تحلیل مؤلفه اصلی دمای خاک، شبکه عصبی مصنوعی، ماشین بردار پشتیبان.

مقدمه

با توجه به اینکه پیچیدگی سیستم‌های هیدرولوژیکی از نظر تعیین تمام پارامترها دخیل و نقص اطلاعات آماری، امکان مدل‌سازی کامل سیستم‌های هیدرولوژیکی را غیرممکن می‌سازد. در چنین شرایطی استفاده از مدل‌سازی سیستمی که مبتنی بر روابط ریاضی باشد، مورد توجه قرار می‌گیرد (کیشی و همکاران، ۲۰۱۶).

از مطالعات صورت گرفته در این زمینه می‌توان به موارد زیر اشاره نمود. در تحقیق توسط زارع ایبانه و همکاران بر روی چند نمونه خاک مختلف، دمای اعماق مختلف خاک در ایستگاه سینوپتیک گرگان توسط روش‌های مختلف هوش مصنوعی تخمین زده شد، که در این میان روش شبکه عصبی مصنوعی توانست نسبت به سایر روش‌ها با دقت بالاتری دمای اعماق مختلف خاک را تخمین بزند (زارع ایبانه و همکاران ۲۰۱۶). سیفی و همکاران (۱۳۹۲)، تحقیقی در زمینه معرفی و کاربرد

دمای خاک یکی از مهم‌ترین پارامترهایی است که متأثر از متغیرهای زیادی همچون بارندگی، سرعت باد، ساعت آفتابی، رطوبت نسبی و ... می‌باشد. (سینگ و ژو، ۱۹۹۷). تخمین دمای خاک در همه مناطق که دارای منابع آب سطحی می‌باشند، به دلیل تأثیرگذاری آن بر روی دریاچه سد، مدل‌سازی بارش-رواناب، عملکرد حوضه رودخانه، مدیریت آب-آبیاری برای محاسبه میزان آب مورد نیاز گیاه و برنامه ریزی آن و ... بسیار مهم است. تخمین دمای خاک ابزار مناسبی برای تعیین تاریخ کشت و کنترل جوانه زنی گیاهان است که در بخش کشاورزی دارای اهمیت فراوان است (حسین زاده طلایی ۲۰۱۴). دمای خاک می‌تواند بطور قابل توجهی در بیلان آب مربوط به رودخانه یا مخزن اثر بگذارد و شاید در پایین آمدن سطح آب نیز مؤثر باشد (مسیکوئن، ۱۹۹۸).

استفاده از پارامترهای تعیین شده توسط سه روش ذکر شده نمود. نتایج نشان داد مدل SVM قابلیت بالایی در تخمین عمق آبشستگی دارد. هوشنگی و همکاران (۱۳۹۴) در تحقیقی بهینه‌سازی تعداد پیژومترها در پیش‌بینی سطح آب‌های زیرزمینی باروش‌های PCA و زمین آمار را مورد بررسی قرار دادند. آن‌ها در مرحله دوم با دو روش خطای اعتبار سنجی متقابل نقاط و تحلیل مولفه اصلی، تعدادی از نقاط را از چرخه تشکیل سطح، خارج و در نهایت با توجه به RMSE سطوح ایجاد شده عملکرد دو روش را با هم مقایسه نمودند. بررسی‌ها نشان داد که روش PCA در مقایسه با روش خطای اعتبار سنجی متقابل از اعتبار بیشتری برخوردار بوده است. تزال و بویکیلدیز (۲۰۱۵) با استفاده از شبکه عصبی مصنوعی و ماشین بردار پشتیبان دمای خاک ماهانه ایستگاه هواشناسی بهشهر را پیش‌بینی نمودند. در این تحقیق از داده‌های دما، رطوبت نسبی، سرعت باد و بارندگی در دوره آماری ۱۹۷۲ تا ۲۰۰۵ استفاده شده است. آن‌ها از مدل ماشین بردار پشتیبان با هسته تابع پایه شعاعی و از شبکه عصبی پرسپترون چند لایه استفاده نمودند. نتایج نشان‌دهنده عملکرد مناسب مدل شبکه عصبی پرسپترون چند لایه با الگوریتم آموزش گرادینان مزدوج کوچک بوده است. نوری و همکاران (۲۰۱۰) روش آزمون گاما و تحلیل مؤلفه‌های اصلی را بر روی عملکرد شبکه عصبی مصنوعی در پیش‌بینی تولید پسماند جامد هفتگی در شهر مشهد را مورد ارزیابی قرار دادند. هدف آن‌ها توسعه یک مدل شبکه عصبی مصنوعی با ساختار ساده و دقت مناسب بوده است. در این تحقیق از روش تحلیل مؤلفه‌های اصلی و آزمون گاما برای کاهش تعداد متغیرهای ورودی استفاده شده است. نتایج مقایسه مدل‌های شبکه عصبی مصنوعی و شبکه عصبی مصنوعی-تحلیل مؤلفه اصلی و شبکه عصبی مصنوعی-آزمون گاما نشان داد که شبکه عصبی مصنوعی-تحلیل مؤلفه اصلی و شبکه عصبی-آزمون گاما نسبت به مدل شبکه عصبی مصنوعی مؤثرتر بوده‌اند.

با توجه به مطالعات صورت گرفته تاکنون مقایسه‌ای بین مدل‌های شبکه عصبی مصنوعی و ماشین بردار پشتیبان همراه با پیش پردازش داده‌ها باروش‌های ذکر شده، صورت نگرفته

حداقل مربعات ماشین بردار پشتیبان در برآورد دمای خاک، تبخیر و تعرق مرجع و تحلیل عدم قطعیت نتایج آن در محدوده شهر کرمان انجام دادند. در این تحقیق از روش حداقل مربعات ماشین بردار پشتیبان مبتنی بر آزمون گاما با شبکه عصبی تابع پایه شعاعی، خطی و چندجمله‌ای برای محاسبه دمای خاک-عرق لایسیمیتری استفاده شده است. نتایج با دومدل شبکه عصبی مصنوعی، سیستم استنتاجی تطبیقی-عصبی فازی و داده‌های لایسیمیتری مقایسه گردید. نتایج حداقل مربعات ماشین بردار پشتیبان بیانگر برتری شبکه عصبی تابع پایه شعاعی نسبت به سایر مدل‌ها بوده است. همچنین تحلیل عدم قطعیت در این تحقیق نشان داد که مدل حداقل مربعات ماشین بردار پشتیبان عدم قطعیت کمتری نسبت به مدل‌های دیگر دارد. کیسی و همکاران در سال ۲۰۱۶ اعماق مختلف خاک را توسط شبکه عصبی مصنوعی و برنامه ریزی ژنتیک در کشور ترکیه مدل سازی کردند و به این نتیجه رسیدند که شبکه عصبی مصنوعی ابزاری مناسب برای تخمین دمای خاک است (کیسی و همکاران ۲۰۱۶). شیخ‌الاسلامی و همکاران (۱۳۹۳) دمای خاک و تعرق گیاه مرجع را با استفاده از روش آنالیز مؤلفه‌های اصلی (PCA) و توسعه مدل رگرسیونی خطی چندگانه در استان مشهد بررسی کردند. در این تحقیق از داده‌های ایستگاه سینوپتیک مشهد در مقیاس روزانه و در دوره آماری ۲۰۰۵-۱۹۹۱ استفاده شده است. آن‌ها تاثیر نسبی پارامترهای دما، رطوبت نسبی، ساعت آفتابی و سرعت باد در ارتفاع دومتری از سطح زمین را بر این پدیده با دومدل رگرسیون خطی چندگانه و رگرسیون خطی چندگانه-تحلیل مؤلفه اصلی مورد بررسی قرار دادند. نتایج نشان داد که مدل رگرسیون خطی چندگانه-تحلیل مؤلفه اصلی دارای عملکرد مناسب‌تری نسبت به مدل رگرسیون خطی چندگانه می‌باشد. سیدیان (۱۳۹۴) از روش‌های نوین در تعیین پارامترهای موثر بر آبشستگی پایه پل استفاده کرد. او با استفاده از سه روش آزمون گاما، تحلیل مؤلفه اصلی و رگرسیون گاما به گام پارامترهای موثر بر آبشستگی را از بین کل پارامترها انتخاب کرده است. سپس با استفاده از مدل ماشین بردار پشتیبان (SVM) اقدام به شبیه سازی و تخمین عمق آبشستگی با

محسوب می شود به همین دلیل مطالعه سیستماتیک منابع آب و خاک، اعم از سطحی و زیرزمینی را ایجاب نموده است. در این مطالعه از داده‌های ایستگاه سینوپتیک تبریز (شکل ۱) استفاده شده است. دوره آماری داده‌ها برای ایستگاه سینوپتیک تبریز از سال ۱۳۹۲ تا ۱۳۹۴ می‌باشد. داده‌های اخذ شده شامل دمای خاک، سرعت باد در ارتفاع دو متری، متوسط دمای هوا، متوسط رطوبت هوا و ساعت آفتابی بوده است. جدول ۱ خصوصیات آماری داده‌های ایستگاه سینوپتیک مورد مطالعه را نشان می‌دهد.

است. هدف از این تحقیق تخمین دمای خاک با استفاده از مدل‌های هوشمند شبکه عصبی مصنوعی و ماشین بردار پشتیبان همراه با کاهش پیچیدگی مدل‌سازی به وسیله پیش پردازش داده‌ها با روش‌های تحلیل مولفه اصلی و آزمون گاما می‌باشد.

مواد و روش‌ها

داده‌ها و منطقه مورد مطالعه

استان آذربایجان شرقی به مرکز تبریز در شمال غربی کشور قرار گرفته است. این منطقه یکی از قطب‌های کشاورزی کشور



شکل ۱- منطقه و ایستگاه‌های مورد مطالعه

جدول ۳-۳- مشخصات آماری ایستگاه سینوپتیک تبریز

مشخصات آماری						متغیرها
انحراف معیار	بیشینه	کمینه	میانگین	کشیدگی	چولگی	
۱۰/۲۶	۳۲/۲	-۱۳	۱۳/۵۸	-۰/۹	-۰/۲۱	میانگین دمای هوا (°c)
۳/۷۴	۱۳/۵	۰	۸/۰۱	-۰/۴۵	-۰/۷۸	میانگین ساعات آفتابی (hr)
۱۸/۲۱	۹۴/۳۷	۱۴/۱۲	۵۱/۱۸	-۱/۰۳	۰/۲۳	میانگین رطوبت نسبی (%)
۱/۷۱	۱۰/۶۲	۰	۳/۷	۰/۰۱۲	۰/۷۳	میانگین سرعت باد (km.hr-1)
۱۱/۷	۳۴/۹۳	-۸/۵۷	۱۶/۵۱	-۱/۳۲	-۰/۰۴	دمای عمق ۱۰ cm (°c)

تحلیل مؤلفه‌های اصلی

که در آن f یک تابع هموار و Γ متغیر تصادفی نشان‌دهنده خطا بوده، همچنین فرض شده است میانگین توزیعی که Γ از آن پیروی کرده، صفر و واریانس خطا، $\text{var}(\Gamma)$ ، کران‌دار است. به این ترتیب، مدل نشان داده‌شده با رابطه ۲ به مدلی هموار با مشتقات جزئی مرتبه اول محدود می‌گردد و آماره گاما (Γ) ، واریانس بخشی از داده‌ها که تابع f قادر به توجیه آن نیست را نشان می‌دهد.

برای بردار ورودی \mathbf{x}_i ، مجموعه $N[i, k]$ ($1 \leq k \leq p$) به نام مجموعه نزدیک‌ترین همسایگی است. پایه و اساس آزمون گاما بر مبنای مجموعه $N[i, k]$ می‌باشد. $\mathbf{x}_{N[i, k]}$ به عنوان k امین نزدیک‌ترین همسایگی‌ها برای بردار مشاهداتی ورودی \mathbf{x}_i تعریف شده به طوری که $(1 \leq k \leq p)$ و $(1 \leq i \leq M)$ می‌باشد. P حداکثر تعداد همسایگی‌ها است که معمولاً بین ۱۰ تا ۵۰ در نظر گرفته می‌شود (دورانت، ۲۰۰۱). به منظور به دست آوردن مقدار آماره گاما (Γ) ، ابتدا مقدار $(\delta_M(\mathbf{k}))$ ، با توجه به داده‌ی ورودی به شرح زیر محاسبه می‌شود.

(۳)

$$\delta_M(\mathbf{k}) = \frac{1}{M} \sum_{i=1}^M |\mathbf{x}_{N[i, k]} - \mathbf{x}_i|^2, 1 \leq k \leq p$$

که در آن نماد $|\dots|$ به عنوان فاصله اقلیدسی بردار $\mathbf{x}_{N[i, k]}$ از بردار همسایگی‌های آن می‌باشد. همچنین مقدار آماره‌ی گامای کوچک، $(\gamma_M(\mathbf{k}))$ با استفاده از داده‌های خروجی به شرح زیر بدست می‌آید.

(۴)

$$\gamma_M(\mathbf{k}) = \frac{1}{2M} \sum_{i=1}^M |\mathbf{y}_{N[i, k]} - \mathbf{y}_i|^2, 1 \leq k \leq p$$

که در آن $\mathbf{y}_{N[i, k]}$ مقدار خروجی نظیر k امین نزدیک‌ترین همسایگی برای بردار \mathbf{x}_i می‌باشد. به این ترتیب، p مقدار برای $\delta_M(\mathbf{k})$ و p مقدار برای $\gamma_M(\mathbf{k})$ به دست آورده می‌شود.

شبکه عصبی پیشرو با آموزش لوانبرگ-مارکورات

از اوایل دهه ۱۹۹۰ تاکنون، این شبکه‌های عصبی مصنوعی در زمینه هیدرولوژی نیز به طور موفقیت‌آمیز به کار رفته گرفته شده‌اند (داوسون و همکاران، ۲۰۰۶). حدود ۹۰٪ شبکه‌هایی

تحلیل مؤلفه‌های اصلی یکی از روش‌های آماری چند متغیره می‌باشد. می‌توان از تحلیل مؤلفه‌های اصلی هنگامی که حجم وسیعی از اطلاعات در اختیار است، برای کاهش پیچیدگی متغیرهای ورودی استفاده کرد (کامدویرن و همکاران، ۲۰۰۵). تحلیل مؤلفه‌های اصلی متغیرهای ورودی را به مؤلفه‌های اصلی که ترکیب مستقل و خطی از متغیرهای ورودی هستند تغییر می‌دهد (لو و همکاران، ۲۰۰۳). به جای استفاده مستقیم از متغیرهای ورودی، می‌توان آنها را به مؤلفه‌های اصلی تبدیل و سپس از آنها همانند متغیرهای ورودی استفاده کرد. در این روش، اطلاعات متغیرهای ورودی با حداقل تلفات در مؤلفه‌های اصلی ارائه خواهد شد (هلنا و همکاران، ۲۰۰۰). برای دسترسی به اطلاعات بیشتر می‌توان به منابع (واکرناگل، ۱۹۹۵؛ تبکینیک و فیدل، ۲۰۰۱؛ نوری و همکاران ۲۰۰۷) مراجعه کرد. در این تحقیق از مجموعه مؤلفه‌های اصلی اولی که حداقل نود و پنج درصد واریانس کل داده‌ها را دارا هستند، به عنوان ورودی مدل‌های تخمین‌گر استفاده شده است.

آزمون گاما

آزمون گاما یک روش غیرمتغیری است که نتایج آن صرف نظر از روش‌های خاص برای مدل‌سازی به کار برده می‌شود (شریفی و همکاران ۱۳۹۲). فرض می‌شود مجموعه داده‌های مشاهداتی به صورت زیر است.

$$\{(\mathbf{x}_i - \mathbf{y}_i), 1 \leq i \leq M\} \quad (1)$$

M

که در آن \mathbf{x}_i بردار مشاهداتی ورودی (مانند رواناب، بارش، درجه حرارت هوا با تأخیرهای مختلف و هر متغیر دیگر موثر)، \mathbf{y}_i خروجی مشاهداتی (مانند میانگین دبی رواناب روزانه) و M تعداد کل مشاهدات می‌باشد.

از بردار \mathbf{x}_i (در فضای m بعدی) برای مدل‌سازی \mathbf{y}_i (در فضای R) استفاده می‌شود. بنابراین، با توجه به فرض فوق رابطه‌ی بین ورودی و خروجی به شرح زیر در نظر گرفته می‌شود.

$$\mathbf{y} = f(\mathbf{x}) + \Gamma \quad (2)$$

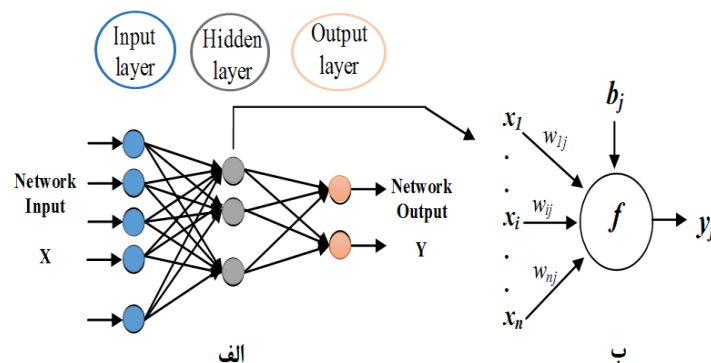
لازم به ذکر است که در تحقیق حاضر از تعداد نرون‌های ۱ تا ۲۰ برای لایه پنهان شبکه استفاده گردید تا از این طریق تأثیر ساختار مختلف شبکه بر عملکرد آن در شبیه‌سازی مقادیر کیفی آبخوان را مورد ارزیابی قرار داد. از آنجا که متداول‌ترین نوع توابع محرک تابع تانژانت سیگموئید می‌باشد بنابراین در این مطالعه از آن برای نگاشت اطلاعات از لایه ورودی به لایه پنهان و همچنین برای نگاشت اطلاعات از لایه پنهان به لایه خروجی استفاده شد.

ماشین بردار پشتیبان

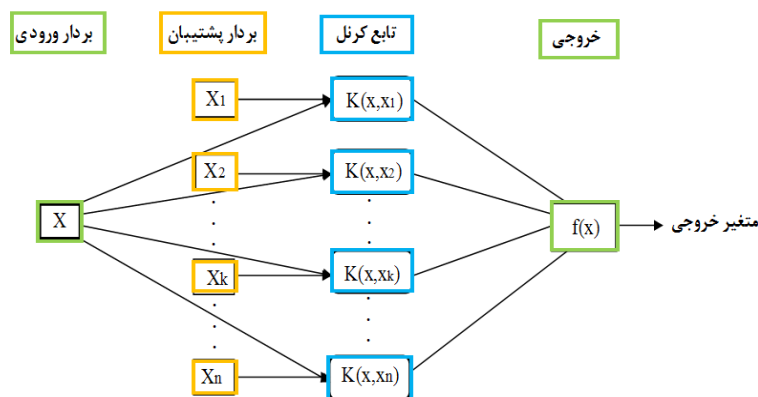
اولین کاربرد این روش در مسائل آب توسط دیباک و همکاران (۲۰۰۱) با شبیه‌سازی بارش- رواناب ارائه شد. ماشین بردار پشتیبان یک سیستم یادگیری کارآمد بر مبنای تئوری بهینه‌سازی مقید است که از اصل استقرای کمینه‌سازی خطای ساختاری استفاده کرده و منجر به یک جواب بهینه کلی می‌گردد. شکل (۳) ساختار ماشین بردار پشتیبان را نشان می‌دهد.

که در مسائل هیدرولوژی کاربرد دارد شبکه‌های پیشرو هستند (کولیالی و همکاران، ۲۰۰۰). معماری معمول این شبکه متشکل از سه لایه ورودی، پنهانی و خروجی است. تعداد نرون‌های موجود در لایه ورودی و خروجی بستگی به نوع مسئله دارد ولی تعداد گره‌های لایه پنهان با سعی و خطا به دست می‌آید.

در شکل ۲، $X = (x_1, \dots, x_i, i=1, \dots, n)$ بردار ورودی‌ها، w_{ij} وزن اتصالی از i امین گره به j امین گره در لایه بعد، b_j بایاس مربوط به گره j ام، Y خروجی نهایی شبکه عصبی مصنوعی است. شبکه‌های پیشرو، اتصالات میان گره‌هایی وجود دارد که در لایه‌های متفاوت قرار دارند. همچنین ورودی‌ها در لایه ورودی شبکه نمایش داده شده و عمل تحریک از ورودی به سمت خروجی شبکه شروع می‌شود (آی اس سی ای، ۲۰۰۰).



شکل ۲- الف) شمای کلی شبکه‌های پیشرو سه لایه، ب) شمای زام گره از آن



شکل ۳- ساختار ماشین بردار پشتیبان

لازم به ذکر است فرآیند محاسبات ماشین بردار پشتیبان، بر اساس کدنویسی در محیط متلب انجام و پارامترهای تابع کرنل از طریق سعی و خطا بهینه شدند.

معیارهای ارزیابی

با استفاده از روش‌های مختلفی می‌توان عملکرد مدل‌ها را مورد ارزیابی و مقایسه قرار داد، یکی از این روش‌ها استفاده از معیارهای ارزیابی می‌باشد. از جمله معیارهای ارزیابی پر کاربرد در علوم مهندسی آب، ضرایب همبستگی (CC)، ریشه میانگین مربعات خطا (RMSE) و خطای مطلق میانگین (MAE) می‌باشند که به ترتیب، در روابط ۱۱، ۱۲ و ۱۳ آورده شده‌اند. دقیق‌ترین مدل با توجه به این معیارها، مدلی خواهد بود که مقدار این سه معیار برای آن به ترتیب، نزدیک به یک، صفر و صفر باشد.

CC

$$CC = \frac{\sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^N (x_i - \bar{x})^2 \cdot \sum_{i=1}^N (y_i - \bar{y})^2}} \quad (11)$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - y_i)^2} \quad (12)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |(y_i) - (x_i)| \quad (13)$$

در روابط بالا \bar{x} و \bar{y} به ترتیب، مقادیر مشاهداتی دمای خاک و میانگین مقادیر مشاهداتی، y_i مقادیر محاسباتی دمای خاک، \bar{y} میانگین مقادیر محاسباتی و N تعداد داده‌ها می‌باشد.

بحث و نتایج

ابتدا برای بررسی اینکه داده‌ها از توزیع نرمال پیروی می‌کنند یا خیر از آزمون کولموگروف-اسمیرنوف استفاده شد. نتایج نشان داد که داده‌ها از توزیع نرمال پیروی نمی‌کنند. همچنین تصادفی بودن داده‌ها با استفاده از آزمون ران تست مورد بررسی قرار گرفت. نتایج مشاهده شده از این آزمون بیانگر تصادفی بودن داده‌ها بوده است.

پس از تعیین ساختار هر یک مدل‌ها و تعیین ترکیب‌های ورودی با دو روش پیش‌پردازش ذکر شده، هر یک از دو مدل

در مدل رگرسیون بردار پشتیبان تابعی مرتبط با متغیر وابسته y که خود تابعی از چند متغیر مستقل x است، برآورد می‌شود. مشابه سایر مسائل رگرسیونی فرض می‌شود رابطه میان متغیرهای مستقل و وابسته با تابع جبری مانند $f(x)$ به علاوه مقداری اغتشاش (خطای مجاز ε) مشخص شود.

$$f(x) = W^T \cdot \phi(x) + b \quad (5)$$

$$y = f(x) + noise \quad (6)$$

چنانچه W بردار ضرایب، b ثابت مشخصه‌های تابع رگرسیونی و $\phi(x)$ نیز تابع کرنل باشد، آنگاه هدف پیدا کردن فرم تابعی برای $f(x)$ است. این مهم با واسنجی مدل ماشین بردار پشتیبان توسط مجموعه‌ای از نمونه‌ها (مجموعه واسنجی) محقق می‌شود. برای محاسبه W و b لازم است تابع خطا (معادله ۳) در مدل ماشین بردار پشتیبان رگرسیونی با در نظر گرفتن شرایط مندرج (قیود) در معادله ۷ بهینه شود.

$$W^T \cdot \phi(x_i) + b - y_i \leq \varepsilon + \varepsilon_i^* \quad (7)$$

$$\frac{1}{2} W^T w + C \sum_{i=1}^N \varepsilon_i + C \sum_{i=1}^N \varepsilon_i^* y_i - W^T \cdot \phi(x_i) - b \leq \varepsilon + \varepsilon_i \quad (8)$$

$$\varepsilon_i, \varepsilon_i^* \geq 0, i = 1, 2, \dots, N$$

در معادلات بالا C عددی صحیح و مثبت است که عامل تعیین جریمه در هنگام رخ دادن خطای واسنجی مدل است، $\phi(x)$ تابع کرنل، N تعداد نمونه‌ها و دو مشخصه ε_i و ε_i^* متغیرهای کمبود هستند (اسکندری و همکاران، ۱۳۹۱). در نهایت، تابع رگرسیون بردار پشتیبان را می‌توان به فرم زیر باز نویسی کرد:

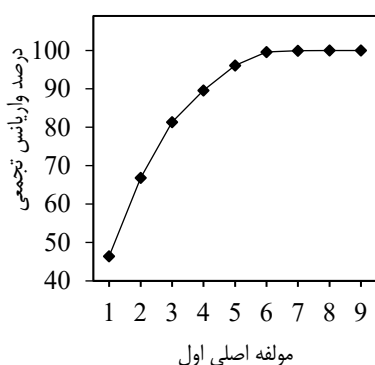
$$f(x) = \sum_{i=1}^N \bar{\alpha}_i \cdot \phi(x_i)^T \cdot \phi(x) + b \quad (9)$$

در معادله (۵)، $\bar{\alpha}_i$ میانگین ضرایب لاگرانژ است. محاسبه $\phi(x)$ در فضای مشخصه آن ممکن است بسیار پیچیده باشد. برای حل این مشکل روند معمول در مدل رگرسیون بردار پشتیبان انتخاب یک تابع کرنل است. با توجه به این که در ماشین بردار پشتیبان معمولاً از تابع کرنل پایه شعاعی استفاده می‌گردد، در این تحقیق نیز از این تابع استفاده گردیده است. فرمول کرنل پایه شعاعی در زیر آورده شده است (کاوزوگلو و کولیکسین، ۲۰۰۹).

$$k(x, x_i) = \exp\left(-\|x - x_i\|^2 / \sigma^2\right) \quad (10)$$

می‌شود. همچنین با توجه به شکل ۴ مولفه ۶،۷،۸ و ۹ برای ایستگاه سینوپتیک تبریز جز ورودی‌های تقریباً بی‌تاثیر می‌باشند و حتی ورود این مولفه‌ها به مدل امکان افزایش خطا را به همراه خواهد داشت.

بنابراین با توجه به شکل ۴ در این تحقیق جهت تخمین دمای خاک ایستگاه سینوپتیک تبریز با استفاده از روش PCA، به ترتیب چهار مؤلفه اصلی اول که درصد بیشتری از واریانس جمعی را تشکیل می‌دادند استفاده شده است، سایر مؤلفه‌های PCA که شامل مؤلفه‌های پنجم تا نهم می‌باشد در مدل سازی تاثیری نداشتند پس در ورودی مدل‌ها لحاظ نشدند.



شکل ۴- درصد واریانس تجمعی مؤلفه‌های اصلی دمای خاک ایستگاه سینوپتیک تبریز

مدل سازی با GT-SVM و GT-ANN

در این بخش از ترکیب ورودی تعیین شده توسط آزمون گاما جهت تخمین دمای خاک ایستگاه‌های مورد نظر استفاده شده است. نتایج تخمین دمای خاک در دوره‌های واسنجی و صحت‌سنجی با استفاده از مدل‌های شبکه عصبی مصنوعی و ماشین بردار پشتیبان در جدول ۳ نشان داده شده است. با توجه به جدول ۳ در ایستگاه تبریز مدل GT-SVM با مجذور میانگین مربعات خطای کمتر نسبت به مدل GT-ANN دارای عملکرد بهتری می‌باشد.

مدل سازی با PCA-SVM و PCA-ANN

نتایج جدول ۴ گویای مدل‌سازی با استفاده از داده‌های حاصل از پیش‌پردازش PCA می‌باشد. مدل‌سازی در این مرحله برای ایستگاه سینوپتیک تبریز با چهار مؤلفه اصلی انجام شده است.

شبکه عصبی مصنوعی و ماشین بردار پشتیبان به‌ازای داده‌های آموزش مورد واسنجی قرار گرفته و سپس عملکرد مدل‌های آموزش دیده به‌ازای داده‌های بخش صحت‌سنجی ارزیابی گردیده است. در ادامه خلاصه نتایج مربوط به هر یک از دو مدل در تخمین دمای خاک ارائه شده است.

نتایج آزمون گاما

برای تعیین پارامترهای موثر جهت تخمین دمای خاک از آزمون استفاده شده است. بنابراین ابتدا آزمون گاما برای کل داده‌ها انجام گرفته (ردیف ۱) و سپس این آزمون براساس حذف هر یک از متغیرها از کل چهار متغیر موجود صورت گرفته است (ردیف‌های ۲ تا ۵) که در جدول ۲ نشان داده شده است. در ایستگاه سینوپتیک تبریز ترکیب‌هایی که پارامترهای دمای متوسط، ساعت آفتابی و سرعت باد از آنها حذف شده است دارای مقدار گامای بیشتری نسبت به حالتی که تمام متغیرها در ترکیب حضور دارند (ترکیب شماره ۱) می‌باشند. بنابراین بایستی از این متغیرها برای تخمین دمای خاک در ایستگاه سینوپتیک تبریز استفاده گردد.

جدول ۲- نتایج آزمون گاما برای ایستگاه سینوپتیک تبریز

تبریز	متغیر غایب	ردیف
گاما		
۰/۰۵۷۱	-	۱
۰/۰۵۷۵	سرعت باد	۲
۰/۰۵۹۱	ساعت آفتابی	۳
۰/۰۶۷۲	دمای متوسط	۴
۰/۰۵۲۵	رطوبت نسبی	۵

تعیین تعداد مؤلفه‌های اصلی ورودی

پس از تعیین مؤلفه‌های اصلی دمای خاک ایستگاه سینوپتیک تبریز، درصد واریانس تجمعی مؤلفه‌های اصلی محاسبه گردیده است. سپس تعداد مؤلفه‌های اصلی اولی که نمود درصد واریانس کل داده‌ها را دارا باشند، تعیین شده‌اند (شکل ۳). با توجه به شکل ۴ برای ایستگاه سینوپتیک تبریز چهار مؤلفه‌های اول ذکر شده موثرترین مؤلفه‌ها برای ورودی مدل‌ها محسوب

با توجه به جدول ۴ مدل PCA-SVM با مجذور میانگین سینوپتیک تبریز عملکرد بهتری در تخمین دمای خاک داشته مربعات خطای کمتر نسبت به مدل PCA-ANN در ایستگاه است.

جدول ۳- نتایج مدل‌سازی با پیش‌پردازش آزمون گاما برای ایستگاه سینوپتیک تبریز

آزمون			آموزش			مدل
CC	MAE(°c)	RMSE(°c)	CC	MAE(°c)	RMSE(°c)	
۰/۹۷	۱/۹۱	۲/۴۸	۰/۹۶	۲/۲۵	۲/۹۲	GT-SVM
۰/۹۷	۲/۰۵	۲/۷۲	۰/۹۶	۲/۰۸	۲/۷	GT-ANN

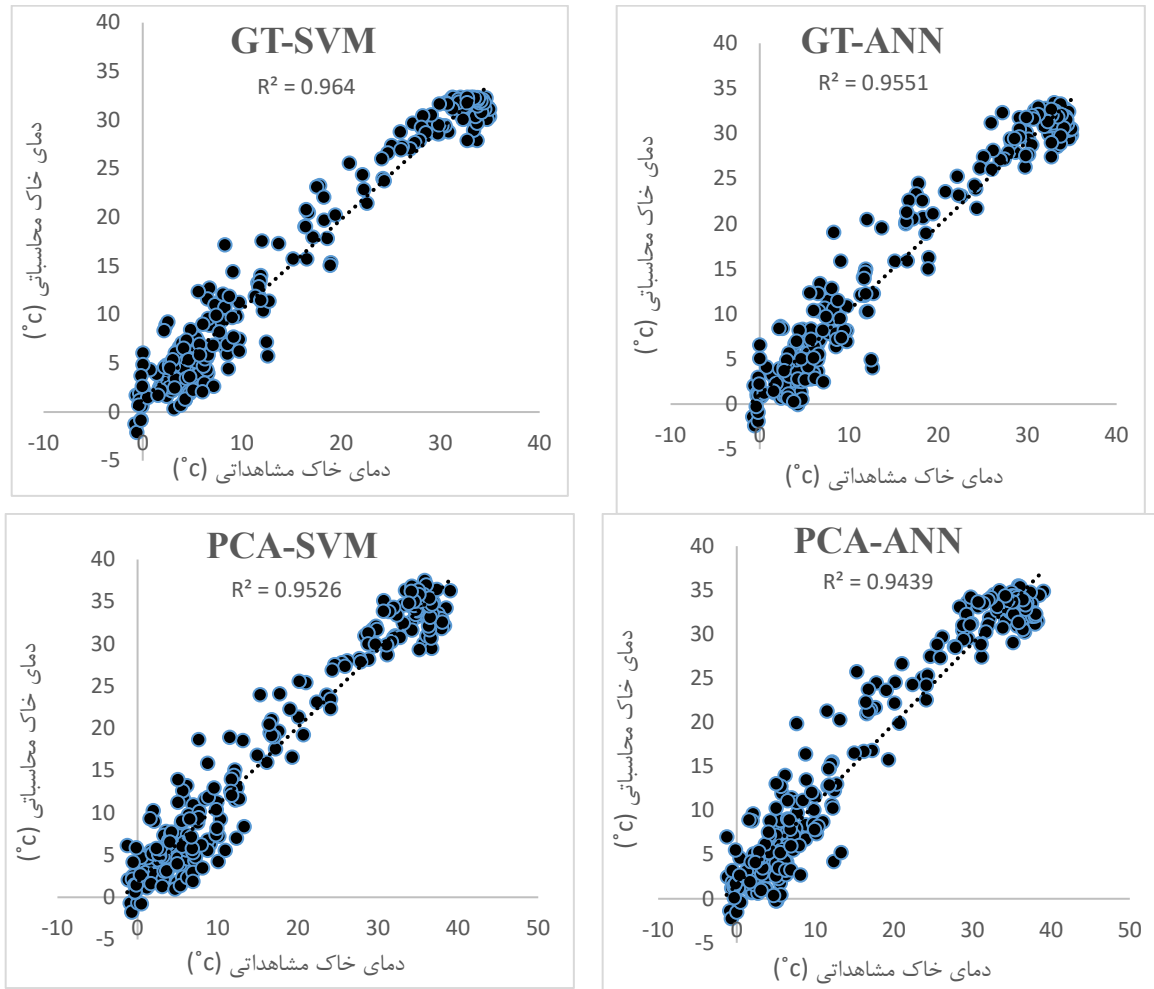
جدول ۴- نتایج مدل‌سازی با پیش‌پردازش تحلیل مؤلفه اصلی برای ایستگاه سینوپتیک تبریز

آزمون			آموزش			مدل
CC	MAE(°c)	RMSE(°c)	CC	MAE(°c)	RMSE(°c)	
۰/۹۷	۲/۳۸	۳/۰۷	۰/۹۵	۲/۵۸	۳/۳۳	PCA-SVM
۰/۹۶	۲/۵۵	۳/۳	۰/۹۶	۲/۲۷	۲/۹۷	PCA-ANN

مقایسه نتایج

نتایج مدل‌های تخمین گر به یکدیگر باشد. بنابراین نمی‌توان یکی از دو روش پیش‌پردازش PCA و گاما را بر دیگری ترجیح داد. به این ترتیب می‌توان نتیجه گرفت که در تخمین دمای خاک ایستگاه سینوپتیک تبریز می‌توان از هر دو روش پیش‌پردازش برای تخمین دمای خاک استفاده کرد. با توجه به نتایج به‌دست آمده چنین می‌توان نتیجه گرفت که مدل‌های PCA-SVM و مدل GT-SVM برای ایستگاه سینوپتیک تبریز عملکرد بهتری نسبت به سایر مدل‌ها داشته‌اند. پس این مدل‌ها به عنوان مدل منتخب برای مدل‌سازی دمای خاک در ایستگاه مذکور استفاده شد و نمودار پراکنش آن‌ها در شکل ۵ نشان داده شده است.

با توجه به نتایج جدول ۴، عملکرد مدل‌های PCA-ANN، GT-ANN، PCA-SVM و GT-SVM در تخمین دمای خاک مورد بررسی بسیار نزدیک به یکدیگر می‌باشند. این شباهت نتایج ناشی از عملکرد روش‌های پیش‌پردازش تحلیل مؤلفه‌های اصلی و آزمون گاما می‌باشد. روش تحلیل مؤلفه‌های اصلی، متغیرهای ورودی را با استفاده از رابطه خطی بین کلیه متغیرهای ورودی به مؤلفه‌های اصلی مستقل از یکدیگر تبدیل می‌کند. در واقع این روش با تعیین ضریب کوچکتر برای متغیرهای ورودی که دارای اطلاعات مشابه هستند، اثر آن متغیر را در مؤلفه‌های اصلی ایجاد شده کم می‌کند. ولی در روش آزمون گاما، با توجه به ضریب گاما بدست آمده از ترکیب‌های ورودی مختلف، متغیری که تاثیر منفی بر روی خروجی تعیین شده دارد، شناسایی گردیده و از ترکیب ورودی نهایی حذف می‌گردد. همانطور که ذکر گردید ماهیت دو روش غیر خطی گاما و خطی PCA با یکدیگر متفاوت هستند. ولی هنگامیکه روش PCA ضریب مربوط به متغیر حذف شده در آزمون گاما را بسیار کوچک در نظر بگیرد، ورودی‌های حاصل از دو روش پیش‌پردازش به یکدیگر نزدیک می‌گردند. این عامل می‌تواند یکی از علل نزدیکی



شکل ۵ - نمودار پراکنش مدل‌های مورد استفاده برای تخمین دمای خاک در ایستگاه سینوپتیک تبریز

نتیجه‌گیری

پارامترهای انتخاب شده ورودی به مدل‌ها توسط آزمون گاما برای ایستگاه سینوپتیک تبریز، شامل دمای متوسط، سرعت باد و ساعت آفتابی بوده است. با توجه به نتایج تحلیل مولفه اصلی بر روی داده‌ها از چهار مولفه اصلی اول برای مدل‌سازی در ایستگاه مذکور استفاده شده است.

نتایج مدل‌سازی در بخش پیش‌پردازش داده‌ها توسط آزمون گاما، نشان می‌دهد که مدل GT-SVM برای ایستگاه سینوپتیک تبریز دارای عملکرد مناسب‌تری بوده اند. و برای بخش تحلیل مولفه اصلی می‌توان در ایستگاه سینوپتیک تبریز از مدل PCA-SVM استفاده نمود.

با توجه جدول‌های ۳ و ۴ و مقایسه دو روش پیش‌پردازش در نحوه عملکرد مدل‌ها، مدل GT-SVM را می‌توان به عنوان مدل منتخب برای ایستگاه سینوپتیک تبریز انتخاب کرد. در

واقع با توجه به عملکرد مدل‌ها می‌توان از هر دو روش پیش‌پردازش برای انتخاب ورودی موثر در تخمین دمای خاک استفاده نمود. همچنین نتایج نشان داد که مدل SVM در تخمین دمای خاک نسبت به مدل ANN با هر دو روش پیش‌پردازش داده‌ها دارای عملکرد مناسب‌تری (با توجه به شاخص‌های آماری ذکر شده) بوده است.

در نهایت با توجه به موارد گفته شده و نتایج حاصله می‌توان چنین نتیجه گرفت که پیش‌پردازش داده‌ها با دو روش ذکر شده، برای تخمین دمای خاک و کاهش پیچیدگی مدل‌ها موفقیت‌آمیز بوده است. به نحوی که با ورودی‌های کمتر و موثرتر باعث شد که دقت مدل‌سازی بیشتر شود و مدت زمان طول فرآیند اجرای مدل‌سازی کاهش پیدا کند. همچنین ورودی‌های اضافی که در تخمین دمای خاک تاثیر کمتری داشتند از ورودی مدل‌ها حذف گردیدند. با انجام این کار

موردی: رودخانه کارون. نشریه آب و فاضلاب. ۳ (۱۸): ۶۰-۶۹.

۷. هوشنگی، ن. آل شیخ، الف و ندیری، ع. الف ۱۳۹۴. بهینه سازی تعداد پیژومترها در ژئوشیمی سطح آب‌های زیرزمینی با روش‌های PCA و زمین آماری. نشریه دانش آب و خاک. ۴/۲ (۲۵): ۵۳-۶۶.

8. ASCE Task Committee on Application of Artificial Neural Networks in Hydrology. 2000. Artificial neural networks in hydrology. I preliminary concepts. Journal of Hydrologic Engineering. 5(2):115-123.

9. Camdevyren, H., Demyr, N., Kanik, A. and S. Keskin. 2005. Use of principal component scores in multiple linear regression models for prediction of Chlorophyll- an in reservoirs. Ecological Modelling. 181(4): 581-589.

10. Coulibaly, P., Anctil, F. and B. Bobée. 2000. Daily reservoir inflow forecasting using artificial neural networks with stopped training approach. Journal of Hydrology. 230(3-4):244-257.

11. Durrant, P.J. 2001. Win_gamma TMA non-linear data analysis and modeling tool with applications to flood prediction. Ph.D. Thesis, Department of Computer Science, Cardiff University Wales, UK.

12. Dibike, Y., Velickov, S., Solomatine, D. and M. Abbott. 2001. Model induction with of support vector machines. Introduction and applications.

هزینه کم‌تری در جمع آوری ورودی‌ها جهت تخمین خاک صورت می‌گیرد. این کار به نوبه خود اهمیت روش‌های پیش پردازش داده‌ها را بیان می‌کند.

منابع

۱. اسکندری، ع. نوری، ر. الف. معراجی، ح و کیاقادی. الف ۱۳۹۱. توسعه مدلی مناسب بر مبنای شبکه عصبی مصنوعی و ماشین بردار پشتیبان برای پیش‌بینی بهنگام اکسیژن‌خواهی بیوشیمیایی ۵ روزه. مجله محیط شناسی. ۶۱ (۳۸): ۷۱-۸۲.

۲. سیفی، ا. میرلطیفی، س.م و ریاحی، ح ۱۳۹۲. معرفی و کاربرد ماشین بردار پشتیبان حداقل مربعات در برآورد دمای خاک- تعرق مرجع و تحلیل عدم قطعیت نتایج، مطالعه موردی شهر کرمان، فصلنامه علمی پژوهشی مهندسی آبیاری و آب. ۱۳ (۴): ۶۷-۷۹.

۳. سیدیان، س.م ۱۳۹۴. استفاده از روش‌های نوین در تعیین پارامترهای موثر بر آبستگي پایه پل. فصلنامه علمی پژوهشی مهندسی آبیاری و آب. ۱۹ (۵): ۱-۱۶.

۴. شریفی، ع. دین‌پژوه، ی. فاخری‌فرد، الف و مقدم‌نیا، ع ۱۳۹۲. ترکیب بهینه متغیرهای برای شبیه‌سازی رواناب در حوزه آبخیز امامه با استفاده از آزمون گاما. زمستان ۱۳۹۲، نشریه دانش آب و خاک. ۴ (۲۳): ۵۹-۷۲.

۵. شیخ‌الاسلامی، ن. قهرمان، ب. مساعدی، ا. داوری، ک و مهاجر پور، م ۱۳۹۳. پیش‌بینی دمای خاک و تعرق گیاه مرجع با استفاده از روش آنالیز مولفه اصلی و توسعه مدل رگرسیون خطی چندگانه، مطالعه موردی: ایستگاه مشهد. نشریه آب و خاک. ۲ (۲۸): ۴۲۰-۴۲۹.

۶. نوری، ر.ا. کراچیان، ر. خدادادی دربان، ا و شکیبایی نیا، ا ۱۳۸۶. ارزیابی اهمیت ایستگاه‌های پایش کیفی رودخانه‌ها با استفاده از آنالیز مولفه‌های اصلی و آنالیز فاکتور، مطالعه

- treeComputers and Electronics in Agriculture. 122: 112-117. <http://dx.doi.org/10.1016/j.compag.2016.01.026>.
19. Lu, W.Z., Wang, W.J., Wang, X.K., Xu, Z.B. and A.Y.T. Leung. 2003. Using improved neural network to analyze RSP, NO_x and NO₂ levels in urban air in Mong Kok, Hong Kong. *Environmental Monitoring and Assessment*. 87(3): 235-254.
20. McCuen, RH. 1998. *Hydrologic analysis and design*. Prentice Hall. EnglewoodCliffs. NewJersey.
21. Noori, R., Karbassi, A. and M. Sabahi. 2010. Evaluation of PCA and gamma test techniques on ANN operation for weekly solid waste prediction. *Journal of Environmental Management*. 91(3): 767-771.
22. Singh, VP. and CY. Xu. 1997. Evaluation and generalization of 13 mass transfer equations for determining free water evaporation. *Hydrological Process*. 11:311-324
23. Tabachnick, B.G. and L.S. Fidell. 2001. *Using Multivariate Statistics*, third ed. Allyn and Bacon, Boston, London.
24. Tezal, G. and M. Buyukyildiz. 2015. Monthly evaporation forecasting using artificial neural networks and support vector machines. *Journal of Computing in Civil Engineering*. 15(3): 208-216.
13. Dawson, C.W., Abrahart, R.J., Shamseldin, A.Y. and R.L. Wibly. 2006. Flood estimation at ungauged sites using artificial neural networks. *Journal of Hydrology*. 319 (1-4): 391-409.
14. Helena, B., Pardo, R., Vega, M., Barrado, E., Fernandez, J.M. and L. Fernandez. 2000. Temporal evolution of groundwater composition in an alluvial aquifer (Pisuerga River, Spain) by principal component analysis. *Water Research* 34(3): 807-816.
15. Hosseinzadeh Talae P. 2014. Daily soil temperature modeling using neuro-fuzzy approach. *Theor. Appl. Climatol*. 118: 481-489,
16. Kavzoglu, T. and I. Colkesen. 2009. A kernel functions analysis for support vector machines for land cover classification. *International Journal of Applied Earth Observation and Geoinformation*. 11(5): 352-359.
17. Kisi, O., Sanikhani, H. Cobaner, M.(2017) Soil temperature modeling at different depths using neuro-fuzzy, neural network, and genetic programming techniques. *Theoretical and Applied Climatology* 129: 833-848.
18. Kisi, O., Genc, O., Dinc, S. and M. Zounemat-Kermani. 2016. Daily pan evaporation modeling using chi-squared automatic interaction detector. *neural networks. Classification and Regression*

26. Zare Abyaneh, H., Bayat Varkeshi, M., Golmohammadi, G. Mohammadi ,K(2016) Soil temperature estimation using an artificial neural network and co-active neuro-fuzzy inference system in two different climates. Arabian Journal of Geosciences.9: 377.
25. Wackernagel, H. 1995. Multivariate Geostatistics: an Introduction with Applications. second ed. Springer. New York and London.