

معرفی و کاربرد ماشین بردار پشتیبان حداقل مربعات در برآورد تبخیر-تعرق مرجع و تحلیل عدم قطعیت نتایج؛ مطالعه موردی شهر کرمان

اکرم سیفی^۱، سید مجید میرلطیفی^۲، حسین ریاحی^۳

تاریخ دریافت: ۹۱/۰۳/۱۰

تاریخ پذیرش: ۹۱/۱۱/۳۰

مقاله مستخرج از پایان نامه کارشناسی ارشد

چکیده

تبخیر-تعرق مرجع (ET_0) یکی از پارامترهای مهم در طراحی پروژه‌های تامین و توزیع آب، مدیریت آبیاری، طراحی سیستم‌های آبیاری، کشاورزی و عملیات هیدرولوژیکی است. پیچیدگی، ناشناخته بودن ریاضیات پدیده تبخیر-تعرق، عدم وجود داده‌های بلندمدت هواشناسی قابل اطمینان، هزینه‌بر بودن استفاده از لایسیترها و عدم وجود آن‌ها در اکثر مناطق لزوم استفاده از روش‌های جدید داده‌کاوی را نشان می‌دهد. بدین منظور در این تحقیق از مدل ماشین بردار پشتیبان حداقل مربعات (LSSVM) مبتنی بر آزمون گاما (GT) با سه تابع هسته‌ای RBF، خطی (Linear) و چند جمله‌ای (Polynomial) برای پیش‌بینی تبخیر-تعرق لایسیتری استفاده گردید و نتایج آن با دو مدل شبکه‌های عصبی مصنوعی (ANN)، سیستم استنتاج تطبیقی عصبی فازی (ANFIS) و داده‌های لایسیتری مقایسه گردید. داده‌های هواشناسی روزانه یکساله ایستگاه سینوپتیک کرمان و داده‌های تبخیر-تعرق لایسیتری در این تحلیل استفاده شد. بهترین ترکیب در مدلسازی ET_0 در ایستگاه مورد بررسی با استفاده از GT، ترکیب دارای متغیرهای دمای حداکثر، دمای نقطه شبنم، رطوبت نسبی میانگین، سرعت باد و شدت تابش انتخاب گردید و مدلسازی بر اساس این ترکیب صورت گرفت. نتایج LSSVM بیانگر برتری تابع هسته‌ای RBF نسبت به دو تابع چندجمله‌ای و خطی بود. علاوه بر این، توزیع خطای پیش‌بینی‌ها نشان داد که مدل‌های ANFIS و LSSVM-RBF میزان خطای کمتری را به ترتیب در دو مرحله آموزش و آزمایشی ایجاد کردند. در انتهای تحقیق، تحلیل عدم قطعیت مونت-کارلو نتایج مدل‌های مختلف مورد استفاده در این تحقیق نیز نشان داد که پیش‌بینی‌های مدل‌های LSSVM عدم قطعیت کمتری نسبت به مدل‌های ANN و ANFIS دارد.

واژه‌های کلیدی: تبخیر-تعرق لایسیتری، تحلیل عدم قطعیت مونت-کارلو، سیستم استنتاج تطبیقی عصبی فازی، شبکه‌های عصبی مصنوعی، ماشین بردار پشتیبان حداقل مربعات.

^۱ دانشجوی دکتری، گروه آبیاری و زهکشی، دانشگاه تربیت مدرس، تهران، ایران، ۰۹۱۳۳۹۰۸۲۹۱، Seifi.akram@gmail.com

^۲ دانشیار، گروه آبیاری و زهکشی، دانشگاه تربیت مدرس، تهران، ایران، ۰۹۱۲۱۰۱۹۳۰۸، m.mirlatifi@gmail.com (نویسنده مسئول)

^۳ استادیار، گروه مهندسی آب، دانشگاه ولی عصر رفسنجان، کرمان، ایران، ۰۹۱۳۴۹۱۰۴۸۱، hossien.riahi@gmail.com

مقدمه

با توجه به ظاهر شدن چهره کم‌آبی، خشکسالی‌ها و نیاز فزاینده بشر به غذای بیشتر، افزایش بهره‌وری آب در زراعت به عنوان اساسی‌ترین مسئله در کشاورزی است که این امر بدون برآورد دقیق نیاز آبی گیاهان میسر نمی‌باشد. در حدود سال ۱۹۷۰، سازمان خواروبار جهانی^۴ (FAO)، روش‌های عملی را برای برآورد نیاز آبی گیاهان توسعه داد که به عنوان استاندارد بویژه برای مطالعات آبیاری پذیرفته شده است. این سازمان معادله پنمن-مانتیث^۵ (PM) را به عنوان روش استاندارد برای برآورد تبخیر-تعرق و اعتبارسنجی معادلات دیگر بیان کرد (آلن و همکاران، ۱۹۹۸). با وجود این تنها تعداد محدودی از ایستگاه‌های هواشناسی، داده‌های هواشناسی قابل اطمینانی را برای کاربرد روش فائو-پنمن-مانتیث (FPM) ثبت می‌نمایند. لزوم استفاده از روش‌های تخمین ET_0 از یک طرف و محدودیت روش‌های موجود به علت نیاز به داده‌های وسیع از طرف دیگر، لزوم استفاده از روش‌های جایگزین را ایجاد می‌کند. در این تحقیق از روش‌های غیرمستقیم شبکه‌های عصبی مصنوعی^۶ (ANN)، سیستم استنتاج تطبیقی عصبی فازی^۷ (ANFIS) و ماشین بردار پشتیبان^۸ (SVM) به منظور برآورد تبخیر-تعرق مرجع^۹ (ET_0) استفاده شده است. SVM یکی از روش‌های یادگیری باناظر است که از آن برای طبقه‌بندی و رگرسیون استفاده می‌شود. رویکرد SVM به این صورت است که در فاز آموزش، سعی می‌شود که مرز تصمیم‌گیری به گونه‌ای انتخاب گردد که حداقل فاصله آن با هر یک از دسته‌های مورد نظر ماکزیمم گردد. این نوع انتخاب باعث می‌شود که تصمیم‌گیری در عمل، شرایط نویزی را به خوبی تحمل کند و پاسخ‌دهی خوبی داشته باشد. SVM روشی برای طبقه‌بندی دوتائی در فضای ویژگی‌های دلخواه است و از این روش مناسب برای مسائل پیش‌بینی بشمار می‌رود (پای و هونگ، ۲۰۰۷).

در این تحقیق مدل SVM برای برآورد ET_0 در ایستگاه لایسمتری کرمان با برنامه‌نویسی توسعه داده شده است و عملکرد آن نسبت به سایر روش‌ها (ANN، ANFIS و روش‌های تجربی و غیرتجربی) سنجیده شد. به منظور توسعه و بررسی عملکرد و عدم قطعیت مدل‌های مختلف از داده‌های یکساله لایسمترهای وزنی کرمان استفاده گردید. از روش آزمون گاما^{۱۰} (GT) به منظور حداقل‌سازی ورودی‌ها و تعیین بهترین پارامترهای ورودی به مدل‌های شبیه‌سازی غیرخطی استفاده گردید. همچنین در این تحقیق از تحلیل عدم قطعیت در بررسی عدم قطعیت مدل‌ها در پیش‌بینی ET_0 استفاده گردید.

از مدل خبره SVM بیشتر در مسائل هیدرولوژیکی و منابع آب استفاده شده است. مدل SVM با استفاده از مسئله بهینه‌سازی دوگانه برای پیش‌بینی رواناب روزانه از رودخانه بختیاری با استفاده از داده‌های هواشناسی و بارندگی استفاده گردید و نتایج آن با مدل ANN- و GA مقایسه شد. روش SVM منجر به مقادیر RMSE کمتری نسبت به دو روش دیگر داشت و روشی سریع‌تر و با دقت بالاتر بود (بهزاد و همکاران، ۲۰۰۹). SVM به عنوان یک روش موفق در مطالعات هیدرولوژیکی مانند پیش‌بینی رواناب (آسفا و همکاران، ۲۰۰۶)، سیلاب (لیانگ و سیوپرگاسم، ۲۰۰۲؛ یو و همکاران، ۲۰۰۶) و کنترل آب زیرزمینی (آسفا و همکاران، ۲۰۰۴) استفاده گردید. در زمینه تبخیر-تعرق نیز توانایی SVM در اصلاح دقت برآورد تبخیر روزانه در مخازن چاه‌نیمه زایل مورد بررسی قرار گرفت (مقدم نیا و همکاران، ۲۰۰۸). اسلامیان و همکاران (۲۰۰۹) برای برآورد تبخیر از گلخانه در مدل SVM و ANN استفاده و نتایج دو مدل را با یکدیگر مقایسه کردند. مقادیر ضرایب همبستگی برای ANN و SVM بترتیب ۰/۹۲ و ۰/۹۶ و مقادیر راندمان بترتیب ۰/۸۳ و ۰/۹۱ بدست آمد که این مقادیر بیانگر توانایی دو مدل SVM و ANN در برآورد تبخیر در گلخانه بود.

آزمون گاما در انتخاب بهترین ترکیبات در پیش‌بینی تبخیر در منطقه چاه‌نیمه زایل با استفاده از ANN (پیری و همکاران، ۲۰۰۹) و ANFIS (مقدم نیا و همکاران، ۲۰۰۹) مورد استفاده قرار گرفت. SVM در ترکیب با

⁴Food and Agriculture Organization

⁵Penman Monteith

⁶Artificial Neural Networks

⁷Adaptive Neuro Fuzzy Inference System

⁸Support Vector Machine

⁹Reference Evapotranspiration

¹⁰ Gamma Test

نزدیکترین همسایه به x است یعنی $|x' - x| > 0$ حداقل می‌باشد. آزمون گاما بر اساس معادله زیر انجام می‌شود.

$$\gamma = \frac{1}{2M} \sum_{i=1}^M (y'(i) - y(i))^2 \quad (1)$$

آماره گاما (Γ) برآوردی از خروجی مدل است که نمی‌تواند بوسیله مدل داده هموار، محاسبه شود. اگر δ مقدار $N[i, p]$ امین نزدیکترین همسایه به $x(i)$ باشد، مقدار δ بصورت زیر تعریف می‌شود (جانز، ۲۰۰۴):

$$\delta(p) = \frac{1}{M} \sum_{i=1}^M \frac{1}{L(N[i, p])} \times \sum_{j \in N[i, p]} |x(i) - x(j)|^2 \quad (2)$$

$$= \frac{1}{M} \sum_{i=1}^M |x(i) - x(N[i, p])|^2$$

که $L(N[i, p])$ طول $N[i, p]$ می‌باشد. بنابراین $\delta(p)$ میانگین مربعات فاصله تا p امین نزدیکترین همسایه می‌باشد (جانز، ۲۰۰۴). همچنین:

$$\gamma_M(p) = \frac{1}{2M} \sum_{i=1}^M \frac{1}{L(N[i, p])} \times \sum_{j \in N[i, p]} (y(i) - y(j))^2 \quad (3)$$

$$= \frac{1}{2M} \sum_{i=1}^M |y(i) - y(N[i, p])|^2$$

که $y(N[i, p])$ مقدار متقابل y برای p امین نزدیکترین همسایه می‌باشد و $1 < p < p_{\max}$ که $p_{\max} \approx 10$. برای محاسبه Γ ، خط رگرسیونی حداقل مربعات برای $(\delta_M(p), \gamma_M(p))$ ایجاد می‌شود (جانز، ۲۰۰۴):

$$\gamma = A\delta + \Gamma \quad (4)$$

عرض از مبدا محور عمودی ($\delta=0$) مقدار Γ می‌باشد. علاوه بر مقدار گاما از مقادیر نسبت V ، آزمون M ، نقاب، شیب و خطای استاندارد برای ارزیابی کیفیت، انتخاب خصوصیات و متغیرهای مهم و تعیین وجود داده‌های کافی برای ساخت یک مدل غیرخطی هموار استفاده می‌شود. اگر مقدار گاما (در مقایسه با واریانس y) بزرگ باشد، احتمال تعیین خروجی با استفاده از مقادیر ورودی کم

روش‌های مدل‌سازی غیرخطی برای برآورد تابش روزانه از مجموعه داده‌های هواشناسی استفاده گردید (رمسن و همکاران، ۲۰۰۸). تحلیل عدم قطعیت بر اساس شبیه‌سازی مونت-کارلو برای مدل‌های ANN و ANFIS در پیش‌بینی ظرفیت کف رودخانه (ریاحی مدوار و ایوب زاده، ۱۳۸۷) و تعیین هندسه رژیم کانال (ریاحی مدوار و همکاران، ۲۰۱۰) انجام گرفت. نتایج نشان داد که مدل ANFIS عدم قطعیت کمتری نسبت به مدل ANN دارد به منظور بررسی قابلیت روش‌های بکار رفته در شبیه‌سازی نوسانات تراز آب شامل روش‌های بیلان آب دریاچه، معادله همبستگی چندگانه و ANN، از تحلیل‌های عدم قطعیت و حساسیت استفاده شد (دلاور و همکاران، ۱۳۸۷). مقادیر ساعتی حاصل از دو مدل F-P-M و پنمن مانیتث انجمن مهندسان آمریکا (ASCE) با مقادیر اندازه‌گیری شده بوسیله لایسیمتر وزنی الکترونیکی کرمان مورد مقایسه قرار گرفتند. برآورد کم مقادیر محاسباتی توسط روش پنمن - مانیتث ASCE نسبت به مقادیر لایسیمتری برای هر یک از ماه‌های مورد مطالعه بدست آمد (بختیاری و همکاران، ۱۳۸۸).

مواد و روش‌ها

آزمون گاما

از آنجایی که تعیین و انتخاب مهم‌ترین و موثرترین پارامترهای یک تابع غیرخطی و ناشناخته در مدل‌های شبیه‌سازی یکی از مشکل‌ترین مراحل توسعه مدل است، در این تحقیق، بهترین ترکیبات از داده‌های ورودی برای ارزیابی تاثیرشان در مدل‌سازی فرآیند ET_0 با استفاده از محاسبه مقدار گاما، تعیین گردید.

GT برآوردی از بهترین میانگین مربعات خطا روی خروجی انجام می‌دهد که بوسیله یک مدل هموار بدست می‌آید. این برآورد، آماره گاما (Γ) نامیده می‌شود. فرض می‌شود که y بوسیله x و یا زیرمجموعه‌ای از مولفه‌های x تعیین شود. برای این منظور مدلی به صورت یک الگوریتم از مجموعه داده‌های اولیه $\{(x_i, y_i), 1 \leq i \leq M\}$ ساخته می‌شود که در آن $x \in R^m$ بردار ورودی است که به تعدادی مجموعه $C \subset R^m$ محدود می‌شود و بطور متقابل اسکالر $y \in R$ مقدار خروجی می‌باشد. دو نمونه داده (x, y) و (x', y') را در نظر گرفته می‌شود که x'

با محدودیت‌های

$$\sum_{i=1}^l (a_i - a_i^*) = 0 \quad (7)$$

$$0 \leq a_i \leq C, i=1,2,\dots,l$$

$$0 \leq a_i^* \leq C, i=1,2,\dots,l$$

اطلاعات کمی در مورد انتخاب تابع غیرخطی مناسب $\Phi(x)$ در دسترس می‌باشد. بعلاوه، محاسبه $\Phi(x_i)$ و $\Phi(x_j)$ در فضای ویژه بسیار پیچیده می‌باشد. محاسبات در فضای ورودی با استفاده از تابع هسته‌ای $k(x_i, x_j) = \langle \phi(x_i), \phi(x_j) \rangle$ انجام می‌شود تا تولید داخلی در فضای ورودی را ایجاد کند. توابع معروف‌ترین توابع هسته‌ای بصورت زیر می‌باشند (یو و همکاران، ۲۰۰۶).

هسته خطی $k(u, v) = u \cdot v$

هسته پلی- $k(u, v) = [(u \cdot v) + c]^d$

نومیال

هسته $k(u, v) = \tanh[\gamma(u \cdot v) + c]$

سیگموئید

هسته تابع پایه $k(u, v) = \exp\left(-\frac{\|u - v\|^2}{2\sigma^2}\right)$

شعاعی

در نهایت، تابع تصمیم رگرسیون بردار پشتیبان غیرخطی بصورت معادله زیر خواهد بود (یو و همکاران، ۲۰۰۶).

$$f(x_i) = \sum_{i=1}^l (-a_i + a_i^*) k(x_i, x_j) + b \quad (8)$$

ماشین بردار پشتیبان حداقل مربعات

سویکنز و وندوال (۱۹۹۹) ماشین بردار پشتیبان با بهینه‌سازی حداقل مربعات^۳ (LSSVM) را معرفی کردند که فرمول‌بندی آن محدودیت مشابهی با SVM را بکار می‌برد. فایده اصلی این روش این است که از نظر محاسباتی بهتر از SVM عمل می‌کند. در این حالت، آموزش نیاز به حل یک مجموعه توابع خطی بجای مسئله برنامه‌نویسی دوگانه که حل کلاسیک SVM می‌باشد، دارد (خمچندانی و همکاران، ۲۰۰۹). این روش بطور موثری پیچیدگی الگوریتم را کاهش می‌دهد. در حالی که در

می‌باشد و در صورتی که گاما کوچک و یا نزدیک به صفر باشد این احتمال افزایش می‌یابد. مقدار V_{ratio} بین صفر تا یک تغییر می‌کند (جانز، ۲۰۰۴). اگر V_{ratio} نزدیک به صفر باشد، به احتمال زیاد داده‌ها از یک تابع هموار f مشتق می‌شود و اگر نزدیک به یک باشد باید هر گونه تلاشی برای مطابقت داده‌ها با مدل هموار را رها کرد (پیری و همکاران، ۲۰۰۹).

ماشین بردار پشتیبان

ماشین بردار پشتیبان (SVM) یکی از روش‌های یادگیری بانظر است که از آن برای طبقه‌بندی و رگرسیون استفاده می‌شود. SVM توسط وپنیک (۱۹۹۸) معرفی شده است و برپایه تئوری یادگیری آماری^۱ بنا نهاده شد. SVM روشی برای طبقه‌بندی دوتائی در فضای ویژگی‌های دلخواه است و از این رو روشی مناسب برای مسائل پیش‌بینی بشمار می‌رود (پای و هونگ، ۲۰۰۷). در بیشتر مسائل جهان واقعی، تابع تقریب خطی استفاده‌های عملی محدودی دارد. وپنیک (۱۹۹۸) با استفاده از مفهوم حاصلضرب داخلی نشان داد که ابتدا می‌توان بردار ورودی x را با یک تبدیل غیرخطی به یک فضای با بعد زیاد انتقال داد و در آن فضا حاصلضرب داخلی را انجام داد و ثابت کرد که اگر یک هسته^۲ متقارن، شرایط قضیهٔ مرسر را داشته باشد، اعمال این هسته در فضای ورودی با بعد کم می‌تواند حاصلضرب را به شدت کاهش دهد. اگر بردار ورودی x_i بوسیله تابع غیرخطی $\Phi(x)$ داخل فضای ویژه نگاشت شود، تابع تصمیم بشکل زیر خواهد بود.

$$f(w, x) = w \cdot \phi(x) + b \quad (5)$$

مسئله رگرسیون غیرخطی می‌تواند مانند مسئله بهینه‌سازی بیان شود:

$$\text{Minimize} \frac{1}{2} \sum_{ij} (a_i - a_i^*)(a_j - a_j^*) \langle \phi(x_i), \phi(x_j) \rangle \quad (6)$$

$$+ \varepsilon \sum_{i=1}^l (a_i + a_i^*) - \sum_{i=1}^l y_i (a_i - a_i^*)$$

¹Statistical Learning Theory

²Kernel

³Least Square Support Vector Machine

رسیدن به نتایج خوب آماری تعداد اعضای نمونه در شبیه‌سازی بین ۵۰۰۰ تا ۲۰۰۰۰ انتخاب شود.

پس از تعیین اندازه نمونه لازم است که متغیرها شبیه‌سازی گردند. برای شبیه‌سازی یک متغیر لازم است که نمونه‌های تولید شده از آن، اولاً از تابع توزیع مربوط به آن پیروی کنند و ثانیاً هر نمونه از نظر آماری مستقل از نمونه‌های دیگر باشد. بدین منظور باید ابتدا تابع توزیع تجمعی متغیر مورد نظر و سپس تابع معکوس توزیع تجمعی احتمال پارامتر مورد نظر مطابق رابطه ۱۰ تعیین گردد.

$$Q(P) = F^{-1}(P) = x_p \quad (10)$$

بدین ترتیب می‌توان با در نظر گرفتن یک احتمال مشخص (P) به مقدار پارامتر مورد نظر (x_p) دسترسی پیدا کرد. در مرحله بعد اعداد تصادفی تولید شده و پس از آن با استفاده از تابع معکوس استخراج شده، می‌توان به مقدار شبیه‌سازی شده پارامتر مورد نظر دسترسی پیدا کرد (دلاور، ۱۳۸۴).

معیارهای ارزیابی

به منظور مقایسه و ارزیابی عملکرد مدل‌های مورد بررسی از پارامترهای میانگین مربعات خطا^۱ (RMSE)، خطای مطلق میانگین^۲ (MAE) و ضریب تعیین (R^2) استفاده گردید. آماره‌های ذکر شده هیچ اطلاعی در مورد نحوه توزیع خطا نمی‌دهند. به این منظور برای ارزیابی مدل‌ها از شاخص میانگین قدرمطلق خطای نسبی^۳ (MARE) و تحلیل آستانه خطا^۴ (TS) نیز استفاده شد و نمودار پراکنندگی خطای مطلق نسبی نیز رسم گردید. شاخص TS_x برای مقدار x درصد از برآوردها، نشان‌دهنده توزیع خطا در مقادیر برآورد شده، برای هر مدل می‌باشد. این شاخص که برحسب درصد تعریف می‌شود برای مقادیر مختلف قدرمطلق خطای نسبی ارائه می‌شود. مقدار TS برای x درصد از برآوردها طبق رابطه ۱۱ و شاخص MARE از رابطه ۱۲ به دست می‌آیند (ریاحی مدوار و ایوب زاده، ۱۳۸۷):

روش SVM از بردارهای پشتیبان برای آموزش و حل مسئله رگرسیونی استفاده می‌شود، در روش LSSVM همه داده‌های آموزشی برای حل مسئله بهینه‌سازی و تولید نتایج استفاده می‌شود. در روش LSSVM داده‌های آموزشی به فضای هسته‌ای نگاشت می‌شوند و برای برقراری توازن بین خطاهای آموزشی و تابع هموار از پارامترهای تنظیم استفاده می‌کند که برای همه نمونه‌ها یکسان است و می‌تواند بعنوان یک پیش‌فرض در نظر گرفته شود. روش LSSVM قادر به حل هر دو مسئله طبقه‌بندی و رگرسیون می‌باشد (والیون و هورواث، ۲۰۰۵). رگرسیون ماشین بردار پشتیبان با حداقل مربعات شامل حل مسئله بهینه‌سازی زیر می‌باشد و بصورت زیر توصیف می‌شود (یو و همکاران، ۲۰۰۶).

$$\text{Minimize}_{\xi, w, b} \frac{C}{2} \|\xi\|^2 + \frac{1}{2} w^2 \quad (9)$$

با محدودیت‌های

$$(w \cdot \varphi(x_i) + b) - y_i + \xi_i = 0 \quad (10)$$

$$i = 1, 2, \dots, M$$

تحلیل عدم قطعیت

مسائل مرتبط با طبیعت همواره شامل دامنه‌ای گسترده از متغیرها و پارامترهای نامعلوم و تصادفی هستند. معمولاً در برآورد و اندازه‌گیری این نوع متغیرها خطاهای زیادی وجود دارد که باعث ایجاد عدم اطمینان در نتایج تحقیقات می‌شوند. بنابراین استفاده از معیاری کمی برای تفسیر و تحلیل میزان عدم قطعیت این مسائل می‌تواند به درک بهتر و قضاوت صحیح کمک کند. در این تحقیق از روش مونت کارلو برای تحلیل عدم قطعیت متغیرهای ایستگاه کرمان (اقلیم بسیار خشک) استفاده گردید که در این راستا روش پیشنهادی مارس و همکاران (۲۰۰۴) مبنا قرار داده شد. اساس کار شبیه‌سازی مونت-کارلو بدین صورت است که متغیرها (ورودی به مدل) با استفاده از تابع توزیع احتمالاتی مربوط بطور تصادفی تولید شده و سپس خروجی معادل آن‌ها از مدل مورد نظر به دست می‌آید و این عمل بارها تکرار می‌شود. سپس عدم قطعیت خروجی (تابع) از طریق محاسبه شاخص‌های آماری و یا تعیین تابع توزیع احتمالاتی برآورد می‌شود (اکهارت و همکاران، ۲۰۰۳). توصیه شده است که برای

¹ Root Mean Square Error

² Mean Absolute Error

³ Mean Absolute Relative Error

⁴ Threshold Statistics

می‌شوند (Allen et al., 1998).

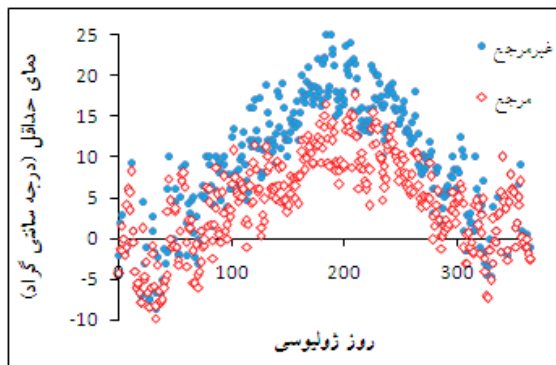
$$\Delta T = (T_{\min})_{obs} - (T_{dew})_{obs} \quad (13)$$

$$(T_{\max})_{cor} = (T_{\max})_{obs} - \left(\frac{\Delta T - K_o}{2} \right) \quad (14)$$

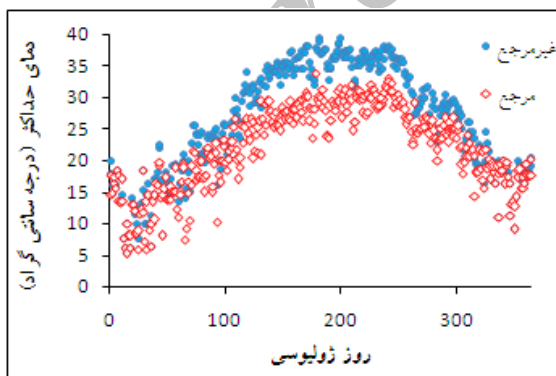
$$(T_{\min})_{cor} = (T_{\min})_{obs} - \left(\frac{\Delta T - K_o}{2} \right) \quad (15)$$

$$(T_{dew})_{cor} = (T_{dew})_{obs} - \left(\frac{\Delta T - K_o}{2} \right) \quad (16)$$

در روابط فوق اندیس obs بیانگر خصوصیت اندازه گیری شده یا محاسبه شده در شرایط غیرمرجع و اندیس cor خصوصیت اصلاح شده نسبت به شرایط مرجع است و K_o ضریب تجربی است که مقدار آن برابر ۲ می‌باشد. شکل‌های ۱ تا ۳ مقادیر دمای حداقل، دمای حداکثر و دمای نقطه شبنم را قبل و بعد از تصحیح نشان می‌دهد.



شکل (۱): نمودار دمای حداقل قبل و بعد از تصحیح



شکل (۲): نمودار دمای حداکثر قبل و بعد از تصحیح

که در آن Y_x تعداد پیش‌بینی شده (از کل تعداد n) برای هر مقدار مطلق خطای نسبی کمتر از x درصد، O_i معادل مقادیر E_{T_o} لایسیمتری و E_i مقادیر برآورد شده می‌باشد (ریاحی مدوار و ایوب زاده، ۱۳۸۷).

$$TS_x = \frac{Y_x}{n} \times 100 \quad (11)$$

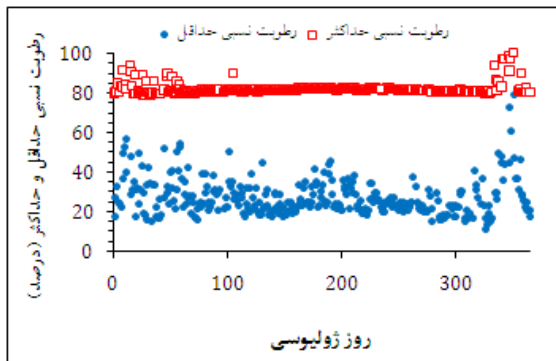
$$MARE = \left| \frac{O_i - E_i}{O_i} \right| \times 100 \quad (12)$$

نتایج و بحث

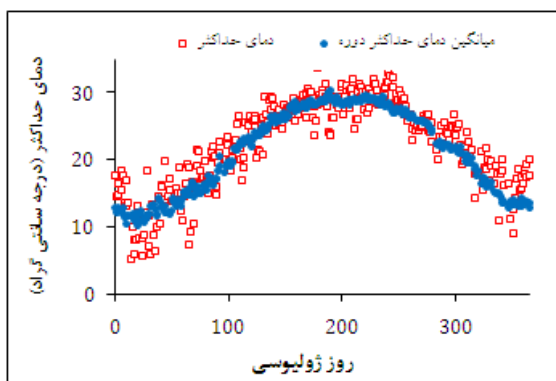
منطقه مورد مطالعه و داده‌های تحقیق

در این تحقیق از داده‌های لایسیمتری دانشگاه شهید باهنر کرمان و داده‌های ایستگاه سینوپتیک کرمان استفاده گردید. دوره داده‌های مورد استفاده در این تحقیق از ماه آگوست سال ۲۰۰۱ تا ماه آگوست سال ۲۰۰۲ می‌باشد که در تحقیق رضایی و همکاران (۱۳۸۶) نیز استفاده شده است. ایستگاه لایسیمتری مورد بررسی در سایت لایسیمتری مزرعه دانشگاه شهید باهنر کرمان با عرض جغرافیایی $30^{\circ}15'N$ ، طول جغرافیایی $58^{\circ}58'E$ و ارتفاع متوسط $1753/8$ متر از سطح دریا واقع شده است. در سطح ایستگاه، گیاه چمن کشت گردیده است (رضایی و همکاران، ۱۳۸۶).

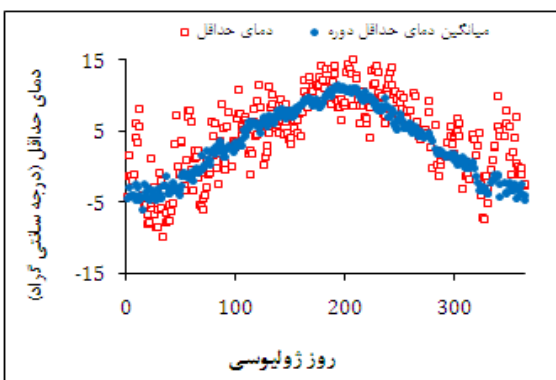
پارامترهای اندازه‌گیری شده در ایستگاه سینوپتیک مورد استفاده در تحقیق حاضر شامل دمای حداقل (T_{\min})، حداکثر (T_{\max})، سرعت باد (u^2)، رطوبت نسبی (RH)، ساعات آفتابی (n) و شدت تابش (R_s) می‌باشند. مقادیر دمای نقطه شبنم (T_{dew}) با استفاده از مقادیر رطوبت نسبی محاسبه گردید. از آنجایی که ایستگاه‌های سینوپتیک در مناطق غیر کشاورزی و بدون پوشش گیاهی (معمولاً در نزدیکی فرودگاه‌ها) واقعند، لذا پارامترهای اندازه‌گیری شده از شرایط مرجع، برای برآورد تبخیر-تعرق برخوردار نیستند و لازم است که قبل از به کار گیری داده‌های این ایستگاه‌ها نسبت به شرایط مرجع تصحیح شوند. روش‌های زیادی برای تصحیح اثر خشکی ایستگاه بر روی داده‌های دما و رطوبت وجود دارد. در این تحقیق از روش پیشنهادی آلن و همکاران (۱۹۹۸) استفاده شد. در این روش مقادیر دمای حداقل، دمای حداکثر و دمای نقطه شبنم (T_{dew}) با استفاده از روابط ۱۳ تا ۱۶ اصلاح



شکل (۵): صحت‌سنجی مقادیر RH ایستگاه سینوپتیک

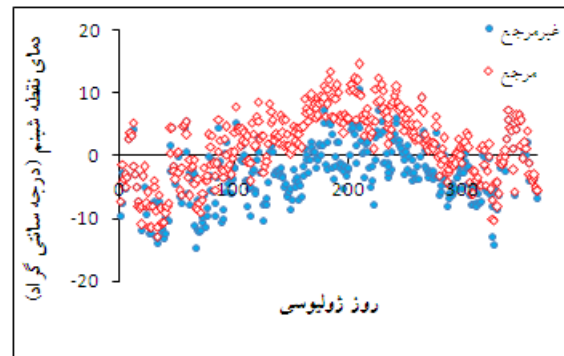


شکل (۶): صحت‌سنجی مقادیر T_{max} ایستگاه سینوپتیک



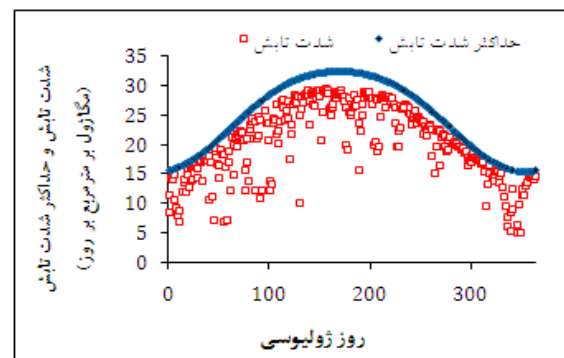
شکل (۷): صحت‌سنجی مقادیر T_{min} ایستگاه سینوپتیک

در این تحقیق از شبکه‌های عصبی پرسپترون چندلایه، استفاده شد. در طراحی ANN طراحی بهینه با یک لایه مخفی صورت گرفت و تعداد بهینه نرون‌های لایه مخفی نیز به روش سعی و خطا به دست آمد. همچنین برای مقادیر اولیه وزن‌ها و بایاس‌ها از مقادیر پیش فرض نرم‌افزار و از تابع فعالیت Logsig-Logsig استفاده شد. همچنین در طراحی شبکه‌های ANFIS از تابع عضویت گوسین استفاده گردید و روش آموزش اصلی در این سیستم، روش پس انتشار خطا می‌باشد. تعداد توابع عضویت برای هر پارامتر از طریق آزمون و خطا تعیین گردید. برای آموزش



شکل (۳): نمودار دمای نقطه شبیم قبل و بعد از تصحیح

صحت‌سنجی مقادیر Rs در شکل ۴ انجام گردید. به دلیل آنکه مقادیر Rs زیر مقادیر حداکثر تابش خورشیدی (RSO) قرار گرفته‌اند کیفیت مناسب داده‌های Rs تایید گردید. با توجه به شکل ۵، مقادیر رطوبت نسبی حداقل بیشتر از ۲۰ درصد هستند که این مقادیر تا ۸۰ درصد نیز رسیده‌اند و همچنین مقادیر رطوبت نسبی حداکثر بیشتر از ۸۰ و کمتر از ۱۰۰ درصد می‌باشند که این اعداد بیانگر کیفیت مناسب داده‌های رطوبت نسبی در دوره مورد نظر می‌باشد. در شکل‌های ۶ و ۷ نمودارهای مربوط به صحت-سنجی مقادیر دمای حداکثر و دمای حداقل آمده است. با توجه به این نمودارها، مقادیر دمای حداکثر و حداقل تقریباً روند قابل قبولی دارند.



شکل (۴): صحت‌سنجی مقادیر Rs ایستگاه سینوپتیک

ملاحظه می‌شود که توابع هسته‌ای RBF و پلی‌نومیال از ترکیب ۴ بهترین نتایج را ایجاد کرد و پیش‌بینی بهتری را نسبت به سایر ترکیبات انجام داد. همچنین مدل‌های ANN و ANFIS با ورودی متناظر با بهترین ترکیب انتخاب شده از مدل LSSVM توسعه داده شدند که نتایج آن در جدول ۳ آمده است. مقایسه نتایج مدل‌های LSSVM با مدل‌های ANN و ANFIS برای این ترکیب نشان می‌دهد که مدل LSSVM-RBF نتایج بهتری ایجاد کرد و رتبه‌بندی مدل‌های مورد استفاده در پیش‌بینی ET_0 برای آن در مرحله آزمون بر اساس معیارهای R^2 ، RMSE و MAE، به ترتیب بصورت LSSVM-RBF، ANFIS، LSSVM-Linear، Polynomial، ANN و LSSVM-Linear بود. در شکل‌های ۹ نمودار توزیع خطای مربوط به مراحل آموزشی و آزمایشی سه مدل مورد استفاده آمده است. مشاهده می‌شود که در مرحله آموزشی در ۹۵ درصد موارد خطای برآورد مدل‌های $LSSVM-RBF_4$ ، ANN_4 و $ANFIS_4$ به ترتیب کمتر از ۳۱، ۳۲ و ۲۸ درصد است و فقط برای ۵ درصد موارد است که قدرمطلق خطای نسبی به ترتیب تا ۵۶، ۵۹ و ۶۷ می‌رسد. در مرحله آزمون نیز در ۹۵ درصد موارد خطای برآورد مدل‌های $LSSVM-RBF_4$ ، ANN_4 و $ANFIS_4$ به ترتیب کمتر از ۲۷، ۲۸ و ۲۹ درصد است و در ۵ درصد موارد قدرمطلق خطای نسبی به ترتیب تا ۴۱، ۶۶ و ۴۷ می‌رسد. بنابراین مدل LSSVM در مرحله آزمون، میزان خطای کمتری را تولید کرده است. در مرحله آزمون قدرمطلق خطای نسبی مدل $ANFIS_4$ کمتر از ANN_4 می‌باشد بنابراین مدل ANFIS نیز در مرحله آزمون خطای کمتری را نسبت به ANN ایجاد کرده است. در شکل ۱۰ روند تغییرات مقادیر ET_0 پیش‌بینی شده با مدل LSSVM و مقادیر لایسیمتری ET_0 و میزان خطای مدل LSSVM آورده شده است. ملاحظه می‌شود که مقادیر ET_0 پیش‌بینی شده با مدل LSSVM از روند داده‌های لایسیمتری تبعیت می‌کند اما پیش‌بینی بخوبی صورت نگرفته و خطای زیادی ایجاد شده است.

مدل‌های مورد استفاده در این تحقیق و همچنین ارزیابی عملکرد روش‌های تجربی تخمین ET_0 در مقابل مقادیر ET_0 لایسیمتری، داده‌ها به دو قسمت تقسیم شد. سری اول، داده‌های آموزشی‌اند که شامل ۷۵ درصد داده‌ها و سری دوم شامل ۲۵ درصد داده‌ها می‌باشد که برای ارزیابی استفاده گردید.

آزمون گاما و انتخاب ترکیب‌های مناسب

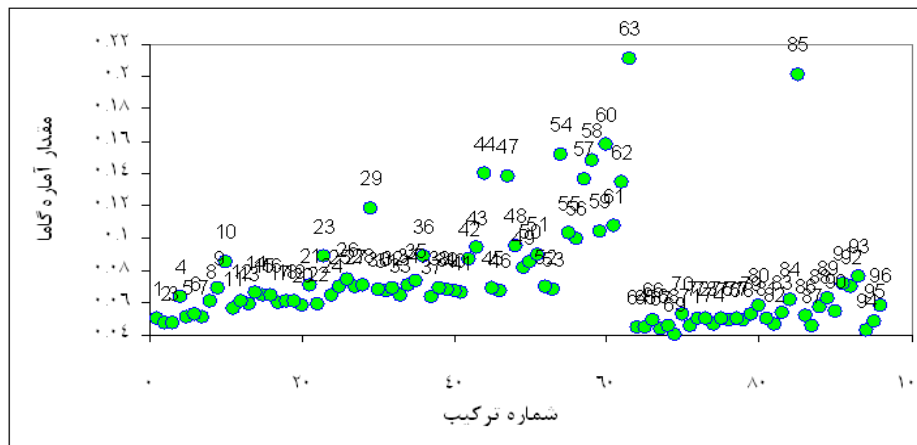
در این تحقیق از ترکیب شش متغیر T_{min} ، T_{max} ، T_{dew} ، RH و u_2 ، n ۶۳ ترکیب متفاوت بدست آمد که شامل ترکیب متغیرهای روش FPM هم بودند، همچنین ۳۰ ترکیب نیز با جایگذاری R_s بجای n ایجاد شد. در نهایت تعداد کل ترکیبات به ۹۳ ترکیب رسید.

نتایج استفاده از آزمون گاما و معیارهای آماری مورد استفاده، در جدول ۱ ارائه شده است. با استفاده از نتایج آزمون گاما سه ترکیب ۳، ۴ و ۵ دارای کمترین مقدار گاما بودند. در این جدول، حداقل مقدار گاما با استفاده از ترکیب شماره ۴ یعنی ترکیب متغیرهای دمای حداکثر، دمای نقطه شبنم، رطوبت نسبی، سرعت باد و شدت تابش دیده می‌شود. ترکیب شماره ۲ دارای بیشترین مقدار گاما بود که این امر آن را بعنوان بدترین ترکیب در پیش‌بینی ET_0 معرفی کرد. مقدار گامای ترکیب شماره ۱ که دارای ورودی یکسان با روش FPM می‌باشد نسبت به ترکیب ۳، ۴ و ۵ بالا می‌باشد که این مطلب اهمیت متغیر شدت تابش در پیش‌بینی ET_0 را نسبت به ساعات آفتابی نشان می‌دهد. در شکل ۸ تمامی مقادیر آماره گامای محاسباتی برای ترکیبات استفاده شده در این تحقیق آورده شد. با توجه به این شکل، ترکیب با ورودی سرعت باد دارای بیشترین مقدار گاما بود که آن را بعنوان بدترین ترکیبات در پیش‌بینی ET_0 معرفی می‌کند.

نتایج ارزیابی مدل LSSVM در مقایسه با داده‌های لایسیمتری برای ترکیب برتر انتخاب شده از طریق آزمون گاما در جدول ۲ آمده است. بر اساس آن تابع هسته‌ای RBF برای تمامی ترکیبات، نتایج بهتری را ایجاد کرد.

جدول (۱): نتایج آزمون گاما در انتخاب بهترین ترکیب پارامترهای ورودی بر مبنای داده‌های لایسیمتری

ترکیب	متغیرهای ورودی	گاما	V _{ratio}
۱	T _{min} , T _{max} , T _{dew} , u ₂ , n	$6/35 \times 10^{-2}$	۰/۲۵۴
۲	u ₂	$21/08 \times 10^{-2}$	۰/۸۴۳
۳	T _{min} , T _{max} , RH, R _s , u ₂	$4/23 \times 10^{-2}$	۰/۱۷۴
۴	T _{max} , T _{dew} , RH, R _s , u ₂	$4/01 \times 10^{-2}$	۰/۱۶۰
۵	T _{min} , T _{max} , R _a , n	$4/34 \times 10^{-2}$	۰/۱۶۹



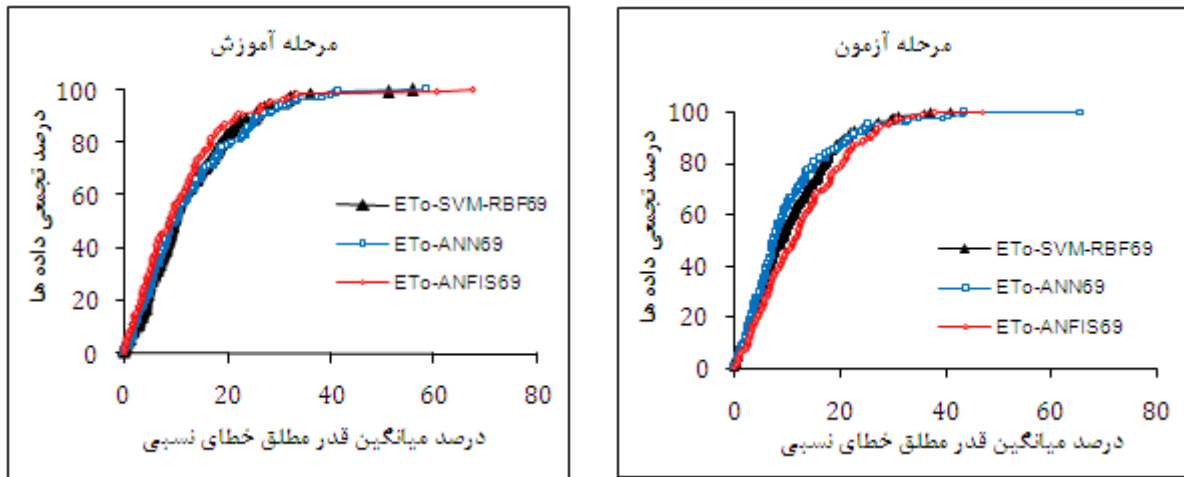
شکل (۸): مقادیر آماره گاما کلیه ترکیبات آزمون شده بر مبنای داده‌های لایسیمتری

جدول (۲): نتایج ارزیابی مدل LSSVM بر مبنای داده‌های لایسیمتری

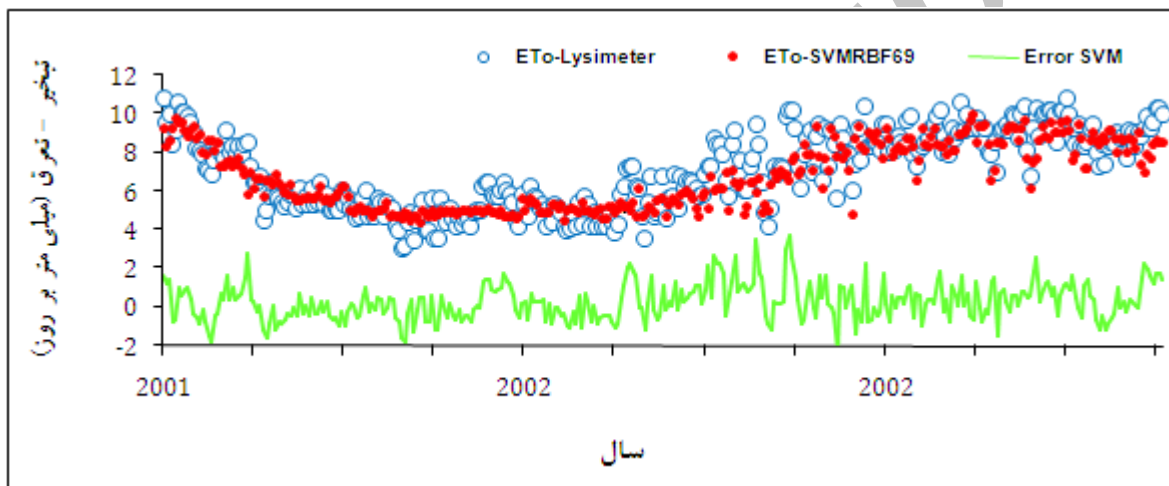
توضیحات	ضرایب بهینه‌سازی	آزمایشی			آموزشی			تابع هسته‌ای	ترکیب
		MAE	RMSE	R ²	MAE	RMSE	R ²		
۲۰/۲۵	۵/۲۳۶	۱/۲۰۰	۱/۵۱۹	۰/۵۱۶	۰/۷۰۹	۰/۸۹۸	۰/۶۸۴	RBF	۱
		۰/۰۷۸	۱/۲۰۲	۱/۵۲۷	۰/۴۷۹	۰/۷۳۰	۰/۹۱۲	پلی نومیال	
		۰/۰۷۷	۱/۴۱۱	۱/۶۵۱	۰/۴۶۴	۰/۹۴۶	۱/۱۳۲	خطی	
۷۶/۰۸	۳۰/۵۵۱	۰/۸۹۶	۱/۱۴۷	۰/۵۴۷	۰/۶۴۲	۰/۸۱۴	۰/۷۴۱	RBF	۳
		۰/۲۷۱	۱/۱۹۳	۱/۴۸۶	۰/۵۴۵	۰/۶۶۸	۰/۸۳۴	پلی نومیال	
		۰/۲۱۵	۰/۸۷۷	۱/۱۲۹	۰/۵۲۴	۰/۸۱۹	۱/۰۱۳	خطی	
۴۱/۷۲	۶/۳۶۵	۰/۸۱۷	۱/۰۶۳	۰/۵۶۵	۰/۵۸۹	۰/۷۳۱	۰/۷۹۱	RBF	۴
		۰/۳۴۷	۰/۸۲۸	۱/۰۶۷	۰/۵۴۸	۰/۶۲۹	۰/۷۸۸	پلی نومیال	
		۱۲/۹۱۵	۱/۴۲۴	۱/۴۱۹	۰/۵۳۰	۰/۷۹۹	۰/۹۸۲	خطی	
۲۵/۰۱	۲۱۰/۰۴	۱/۱۸۸	۱/۵۰۷	۰/۵۱۰	۰/۶۴۹	۰/۸۱۷	۰/۷۳۸	RBF	۵
		۰/۲۰۶	۱/۱۹۹	۱/۵۱۵	۰/۴۸۴	۰/۶۷۸	۰/۸۴۶	پلی نومیال	
		۰/۴۲۶	۱/۳۶۰	۱/۶۰۸	۰/۴۸۱	۰/۸۳۵	۱/۰۱۸	خطی	

جدول (۳): نتایج ارزیابی مدل های ANN و ANFIS بر مبنای داده های لایسیمتری

توضیحات	آزمایشی			آموزشی			شماره ترکیب
	MAE	RMSE	R ²	MAE	RMSE	R ²	
معماری شبکه ۱-۵-۵	۰/۹۵۱	۱/۱۷۵	۰/۵۳۵	۰/۸۵۹	۱/۰۹۲	۰/۶۴۱	ANN ₄
توابع عضویت هر پارامتر ورودی ۴	۱/۲۷۶	۱/۰۴۵	۰/۵۶۴	۰/۷۸۱	۰/۶۰۷	۰/۷۶۱	ANFIS ₄



شکل (۹): پراکندگی و توزیع خطای نسبی در مرحله آموزش و آزمون مدل‌ها

شکل (۱۰): روند تغییرات مقادیر ET_0 پیش‌بینی شده از مدل RBF_4 در مقایسه با داده‌های لایسیمتری

۵۰۰۰ نمونه تولید شده از طریق اعمال ماتریس همبستگی انجام گرفت و معلوم شد که متغیرها مستقل از یکدیگر می‌باشند. پس از تولید نمونه‌های ورودی، خروجی مدل‌های $LSSVM$ ، ANN و $ANFIS$ بر اساس آنها محاسبه و در نهایت مشخصات آماری و نحوه توزیع آنها تعیین گردید. مقادیر بیشتر انحراف معیار، واریانس و محدوده تغییرات مدل‌ها بیانگر وجود عدم قطعیت بیشتر در آنهاست. بر این اساس در جدول (۴) رتبه‌بندی مدل‌ها از لحاظ عدم قطعیت انجام گردید. ملاحظه می‌شود که مقدار عدم قطعیت می‌باشد. همچنین عدم قطعیت مدل‌های $LSSVM$ کمتر از مدل‌های ANN و $ANFIS$ در پیش‌بینی ET_0 می‌باشد.

تحلیل عدم قطعیت مدل‌های پیش‌بینی کننده ET_0
در این قسمت تحلیل عدم قطعیت به منظور مقایسه عدم قطعیت مدل‌های $LSSVM$ ، ANN و $ANFIS$ با بهترین ترکیب انتخاب شده بر اساس داده‌های لایسیمتری در ایستگاه کرمان صورت گرفته است و نتایج آنها با یکدیگر مقایسه شد. برای کاربرد روش مونت-کارلو، بهترین توزیع آماری برای هر متغیر تعیین شد. پس از تعیین تابع توزیع هر متغیر، از طریق الگوریتم‌های نمونه‌گیری به روش کاملاً تصادفی، نمونه‌های لازم با استفاده از معکوس تابع توزیع احتمال هر متغیر براساس ۵۰۰۰ نمونه تصادفی تولید شده در نرم افزار MATLAB تولید گردید. بدین ترتیب، تولید نمونه‌ها بر اساس معادله مرتبط با هر پارامتر و با در نظر گرفتن همبستگی بین

جدول (۴): مشخصات آماری خروجی مدل‌ها بر اساس نمونه‌های تصادفی

مدل‌های مورد بررسی	میانگین (mm/day)	انحراف معیار (mm/day)	واریانس (mm/day)	دامنه تغییرات (mm/day)	رتبه‌بندی کمترین عدم قطعیت
ET _o - LSSVM-LIN ₄	۹/۵۷۴	۰/۹۱۵	۰/۸۳۷	۶/۵۲	۲
فائو-پنمن-مانتیت	۳/۹۷۸	۱/۷۱۷	۲/۹۴۷	۶/۵۵	۱
ET _o - SVM-POLY ₄	۹/۴۳۰	۱/۹۱۲	۳/۶۵۷	۱۴/۶۴	۳
ET _o - SVM-RBF ₄	۷/۰۸۸	۱/۹۷۲	۳/۸۸۹	۶/۸۲	۴
ET _o - ANN ₄	۱۱/۷۸۴	۴/۱۸۴	۱۷/۵۰۵	۱۸/۸۳	۵
ET _o - ANFIS ₄	۱۲/۷۲۲	۵/۶۲۰	۳۱/۵۸۲	۳۱/۴۶	۶

نتیجه‌گیری

به دلیل اهمیت برآورد پدیده تبخیر-تعرق مرجع (ET_o) در سیستم‌های برنامه‌ریزی و مدیریت آبیاری، در این تحقیق از مدل ماشین بردار پشتیبان حداقل مربعات (LSSVM) بعنوان سیستمی خیره برای پیش‌بینی این پدیده استفاده گردید. از سه نوع تابع هسته‌ای RBF، خطی و چندجمله‌ای در سیستم LSSVM استفاده گردید و نتایج آنها با مدل‌های ANN و ANFIS، روش فائو-پنمن-مانتیت و داده‌های لایسیمتری مقایسه گردید. در این تحقیق از ۹۳ ترکیب حاصل از هفت متغیر هواشناسی به منظور برآورد ET_o استفاده گردید. از آنجایی که ورود همه این ترکیبات به مدل‌های برآورد کننده مشکل و وقت‌گیر بود، در این تحقیق از آزمون گاما به منظور تعیین بهترین ترکیب در پیش‌بینی ET_o استفاده گردید. داده‌های یکساله ایستگاه سینوپتیک کرمان متناظر با داده‌های تبخیر-تعرق لایسیمتری بدین منظور استفاده گردید. بر اساس داده‌های

لایسیمتری، ترکیب شامل متغیرهای دمای حداکثر، دمای نقطه شبنم، رطوبت نسبی میانگین، سرعت باد و شدت تابش را بعنوان بهترین ترکیب و ترکیب شامل متغیر سرعت باد را بعنوان بدترین ترکیب معرفی کرد. مدل‌های LSSVM، ANN و ANFIS برای ترکیب انتخاب شده از طریق آزمون گاما اجرا گردیدند که بر اساس نتایج، مدل‌های LSSVM-RBF بهترین نتایج را در مقایسه با مدل‌های ANN و ANFIS ایجاد کردند. همچنین رسم نمودارهای نحوه پراکندگی و توزیع مقادیر مطلق خطای نسبی در دو مرحله آموزشی و آزمایشی حاکی از آن بود که مدل‌های ANFIS و LSSVM-RBF میزان خطای کمتری را به ترتیب در دو مرحله آموزش و آزمایشی ایجاد کردند. نتایج تحلیل عدم قطعیت بر مبنای داده‌های لایسیمتری حاکی از عدم قطعیت کمتر مدل‌های LSSVM نسبت به دو مدل ANN و ANFIS در پیش‌بینی ET_o بود.

منابع

- ۱- بختیاری، ب.، ع. لیاقت، ع. خلیلی و م.ج. خانجانی. ۱۳۸۸. ارزیابی دو مدل ترکیبی برآورد تبخیر-تعرق مرجع چمن در بازه زمانی ساعتی (مطالعه موردی اقلیم کرمان). علوم و فنون کشاورزی و منابع طبیعی، علوم آب و خاک، سال سیزدهم، شماره ۵۰.
- ۲- دلاور، م. ۱۳۸۴. تحلیل و ارائه مدل نوسانات تراز آب دریاچه ارومیه و آنالیز ریسک مناطق ساحلی. پایان نامه کارشناسی ارشد سازه‌های آبی، دانشکده کشاورزی، دانشگاه تربیت مدرس. ۱۴۰ ص.
- ۳- دلاور، م.، ا. مرید و م. شفیع‌فر. ۱۳۸۷. شبیه‌سازی، تحلیل حساسیت و عدم قطعیت تراز آب دریاچه ارومیه نسبت به مولفه‌های بیلان آبی آن. مجله هیدرولیک، جلد ۳، شماره ۱.
- ۴- رضایی، ع.، ب. بختیاری، ف. هوشیاری پور و م. دهقانی امیری. ۱۳۸۶. ارزیابی روشهای مختلف برآورد تبخیر تعرق گیاه مرجع با استفاده از سنجش‌های لایسیمتری (مطالعه موردی: شهر کرمان). نهمین سمینار سراسری آبیاری و کاهش تبخیر، کرمان.
- ۵- ریاحی، ح. و س.ع. ایوب‌زاده. ۱۳۸۷. تخمین ضریب پراکندگی طولی آلودگی با استفاده از سیستم استنتاج فازی-عصبی تطابقی. مجله علمی و پژوهشی آب و فاضلاب، شماره ۶۷، صفحات ۴۶-۳۴.
- 6- Allen, R. G.; L. S. Preira; D. Raes; M. Smith. 1998. Crop evapotranspiration guidelines for computing crop water requirement. FAO Irrigation and Drainage Paper, NO.56, Rome, Italy.

- 7- Asefa, T.; MW. Kemblowski; M. Mckee; A. Khalil. 2006. Multi-time scale stream flow prediction: The support vector machine approach. *Hydrology*, 318:7-16.
- 8- Asefa, T.; MW. Kemblowski; G. Urroz; M. Mckee; A. Khalil. 2004. Support vector-based groundwater head observation networks design. *Water Resource Research*, 40, W11509, doi:10.1029/2004WR003304.
- 9- Behzad, M.; K. Asghari; M. Eazi; M. Pallhang, 2009. Generalization performance of support vector machines and neural networks in runoff modeling. *Expert Systems with Applications*, 36: 7624-7629.
- 10- Eckhardt, K., L. Breuer; H-G. Frede. 2003. Parameter uncertainty and the significance of 12 simulated land use change effects. *Journal of Hydrology*, 273:164-176.
- 11-Eslamian, S.S.; J. Abedi-Koupai; MJ. Amiri; SA. Gohari. 2009. Estimation of daily reference evapotranspiration using support vector machines and artificial neural networks in greenhouse. *Environmental Sciences*, 4: 439-447.
- 12-Jones, A. J. 2004. New tools in non-linear modeling and prediction. *Computational Management Science*, DOI: 10.1007/s10287-003-0006-1, 109-149.
- 13-Khemchandani, R.; Jayadeva; S. Chandra. 2009. Regularized least squares fuzzy support vector machine time series forecasting. *Expert System with Application*, 36: 132-138.
- 14- Liong ,SY.; C. Sivapragasam. 2002. Flood stage forecasting with support vector machines. *American Water Resource Association*, 38:173-186.
- 15- Marce, R.; M. Comerma; J.C. Garcia; J. Armengol. 2004. A neuro-fuzzy modeling tool to estimate fluvial nutrient loads in watersheds under time-varying human impact. *Limnology and Oceanography*, 2:342-355.
- 16-Moghadamnia, A.; M. Ghafari Gousheh; J. Piri; S. Amin; D. Han. 2009. Evaporation estimation using artificial neural network and adaptive neuro-fuzzy inference system technique. *Advance in Water Resource*, 32: 88-97.
- 17-Moghadamnia, A.; M. Ghafari; J. Piri; D. Han,. 2008. Evaporation estimation using support vector machines technique. *Engineering and Technology*, 33:14-22.
- 18- Pai, PF.; WC. Hong. 2007. A recurrent support vector regression model in rainfall forecasting. *Hydrological Process*, 21:819-827.
- 19-Piri, J.; S. Amin; A. Moghadamnia; A. Keshavarz; D. Han; R. Remesan. 2009. Daily pan evaporation modeling in a hot and dry climate. *Hydrologic Engineering*, 14(8): 803-811.
- 20-Remesan, R.; M.A. Shamim; D. Han. 2008. Model data selection using gamma test for daily solar radiation estimation. *Hydrological Processes*, 22:4301-4309.
- 21-Riahi-Madvar, H.; S. A. Ayyoubzadeh. 2010. Uncertainty analysis of ANN and ANFIS technique for predicting bed load transport. *Hydraulic modeling and uncertainty*, 2-4 June, Sophia Antipolis.
- 22-Riahi-Madvar, H.; S. A. Ayyoubzadeh; R. Noori. 2010. Uncertainty analysis of ANN and ANFIS technique in comparison with regime equations for determination of regime channel geometry. *Tenth Symposium on Stochastic Hydraulics, Fifth International Conference on Water Resources and Environment Research*, 5-7 July, Quebec City, Canada.
- 23-Suykens, J.A.K.; J. Vandewalle. 1999. *Least Squares Support Vector Machine Classifiers*. Kluwer Academic Publishers. Printed in the Netherlands.
- 24-Valyon, J.; G. Horvath. 2005. A robust LS-SVM regression. *World Academy of Science, Engineering and Technology*, 7:148-153
- 25-Vapnik, V.N.1998. *Statistical Learning Theory*. Wiley, New York.
- 26- Yu, P.S.; S.T. Chen; I.F. Chang. 2006. Support vector regression for real-time flood stage forecasting. *Hydrology*, 328: 704-716.

Introduction and application of Least Square Support Vector Machine (LSSVM) for simulation of reference evapotranspiration and uncertainty analysis of results, A case study of the Kerman city

A. Seefi, M. Mirlatifi, H. Reahi

Abstract

Estimating reference evapotranspiration (ET_0) is one of the most important variables at water supply and distribution, irrigation management, irrigation systems design, agriculture and hydrological operations. The need for accurate estimates of ET_0 , complexity of ET_0 , unknowing mathematical of phenomenon, lack of reliable meteorological data, the cost of using lysimeters and their absent in most areas magnifies the need for developing new data mining methods. In this paper, the Least Square Support Vector Machine (LSSVM) model with three kernels function of RBF, Linear and Polynomial based on Gamma test used for estimating ET_0 and their results compared with other methods including Artificial Neural Networks (ANN) and Adaptive Neuro Fuzzy Inference System (ANFIS) models. In this analysis, annual daily meteorological data of Kerman synoptic station and ET_0 lysimeter data used. In this research, Gamma test was used for selecting the best combinations of input parameters for various models used instead of using trial and error classic methods. The combination including maximum and dew point temperatures, relative humidity, wind speed, and solar radiation selected as the best combination for estimating ET_0 and modeling was based on this combination. The LSSVM model with RBF kernel performed better than the SVM model with polynomial and linear kernels. Additionally, the distribution of prediction error was calculated that the ANFIS and LSSVM-RBF created less error in train and test steps, respectively. At the end study, Monte-Carlo uncertainty analysis was performed on results of different models that were used in this study. According to the results, predictions of LSSVM models showed less uncertainty than the ANFIS and ANN models were used.

Keywords: Lysimeter Evapotranspiration, Monte-Carlo uncertainty Analysis, Adaptive Neuro Fuzzy Inference System, Artificial Neural Networks, Least Square Support Vector Machine.