

ارزیابی مدل حداقل مربعات ماشین بردار پشتیبان در برآورد تبخیر و مقایسه آن  
با مدل‌های تجربی

محدثه کاوسی، عباس خاشعی سیوکی، محسن پوررضابیلندی و محمد حسین نجفی

دوره ۳، شماره ۳، پاییز ۱۳۹۶، صفحات ۲۳۵-۲۴۷

Vol. 3(3), Autumn 2017, 235 – 247

**Evaluation of LSSVM Model to Estimate  
Evaporation and its Comparison with  
Empirical Models**

Kavoosi M., Khasheisiuki A., Pourrezabilondi M.  
and Najafi M. H.



[www.jewe.ir](http://www.jewe.ir)

OPEN ACCESS

نحوه ارجاع: کاوسی م.، خاشعی سیوکی ع.، پوررضابیلندی م. و نجفی م. ح. (۱۳۹۶). ارزیابی مدل حداقل مربعات ماشین بردار پشتیبان در برآورد تبخیر و مقایسه آن با مدل‌های تجربی. محیط زیست و مهندسی آب، جلد ۳، شماره ۳، صفحات: ۲۳۵- ۲۴۷

**Citing this paper:** Kavoosi M., Khasheisiuki A., Pourrezabilondi M. and Najafi M. H. (2017). Evaluation of LSSVM model to estimate evaporation and its comparison with empirical models. J. Environ. Water Eng., 3(3), 235 - 247.

## ارزیابی مدل حداقل مربعات ماشین بردار پشتیبان در برآورد تبخیر و مقایسه آن با مدل‌های تجربی

محدثه کاوسی<sup>۱</sup>، عباس خاشعی سیوکی<sup>۲\*</sup>، محسن پوررضابیلندی<sup>۳</sup> و محمد حسین نجفی<sup>۳</sup>

<sup>۱</sup> کارشناسی ارشد، گروه علوم و مهندسی آب، دانشکده کشاورزی، دانشگاه بیرجند، بیرجند، ایران

<sup>۲</sup> دانشیار، گروه مهندسی آب، دانشکده کشاورزی، دانشگاه بیرجند، بیرجند، ایران

<sup>۳</sup> استادیار، گروه مهندسی آب، دانشکده کشاورزی، دانشگاه بیرجند، بیرجند، ایران

\* نویسنده مسئول: abbaskhashei@birjand.ac.ir

مقاله اصلی

تاریخ دریافت: [۱۳۹۶/۰۳/۰۹]

تاریخ پذیرش: [۱۳۹۶/۰۶/۲۰]

### چکیده

در این تحقیق با استفاده از پارامترهای هواشناسی در دشت بیرجند در استان خراسان جنوبی در دوره ۱۶ ساله به ارزیابی عملکرد آزمون گاما و مقایسه دقت مدل‌های حداقل مربعات ماشین بردار و روش‌های تجربی به منظور تخمین میزان تبخیر پرداخته شد. با استفاده از روش آزمون گاما از میان پارامترهای تأثیرگذار بر تبخیر، پارامترهای بهینه ورودی جهت مدل‌سازی تخمین تبخیر از میان ۹۰ ترکیب معین، تعیین گردید. تعداد ۷ ترکیب برتر نسبت به ترکیب‌های دیگر به دست آمد، سپس ترکیب بهینه با استفاده از مدل حداقل مربعات ماشین بردار پشتیبان با توابع کرنل مختلف و روش‌های تجربی مورد ارزیابی قرار گرفتند. نتایج نشان‌دهنده عملکرد مدل LSSVM-poly (چندجمله‌ای)، در ترکیب با متغیرهای دمای کمینه، دمای میانگین، سرعت باد و ساعات آفتابی با شاخص‌های عملکرد (RMSE= ۰/۹۹۵ و  $R^2=۰/۸۹۱۵$ ) نسبت به توابع کرنل و دیگر روش‌های تجربی دارای دقت بالاتری است.

واژه‌های کلیدی: آزمون گاما؛ تبخیر؛ LSSVM؛ بیرجند

## ۱- مقدمه

در مهندسی هیدرولوژی، تبخیر از دو نظر حائز اهمیت است؛ اولاً از آنجایی که تبخیر از سطح رودخانه‌ها، دریاچه‌ها و مخازن سدها باعث تلفات آب می‌شود، لازم است مقدار آن محاسبه شود. ثانیاً تبخیر و تعرق از سطح خاک و پوشش گیاهی در داخل حوزه‌های آبریز یکی از اجزای چرخه آب به‌شمار می‌رود (Heidary et al. 2013). عوامل مؤثر بر تبخیر زیاد هستند، از جمله می‌توان به تابش خورشید، سرعت باد، شوری آب و سطح تبخیر اشاره کرد (Varkeshi et al. 2010).

Jahanbakhsh et al. (2001) روش‌های ترکیبی، دمایی، تابشی و همبستگی چندگانه و رطوبتی را برای ایستگاه تبریز با آمار ۲۰ ساله محاسبه و سپس با روش تست تبخیر مقایسه کردند. نتیجه تحقیق آن‌ها نشان داد که روش کریستینسن‌هاگریوز در قیاس با سایر روش‌ها انطباق بیشتری با روش تست تبخیر دارد. Hozhbar et al. (2015) به برآورد تبخیر و تعرق مرجع با استفاده از مدل‌های تجربی و مدل‌سازی آن با شبکه‌عصبی مصنوعی و مقایسه آن با داده‌های لایسیمتری در ایستگاه کهریز ارومیه پرداختند. نتایج حاصل از محاسبات نشان داد که شبکه‌عصبی مصنوعی نسبت به تمامی روش‌های کلاسیک عملکرد مناسب‌تری دارد. همچنین در روش‌های کلاسیک، مدل تورک در اولویت بعدی قرار دارد.

Adl et al. (2009) میزان تبخیر و تعرق گیاه مرجع را با شش روش تجربی و به کمک داده‌های اقلیمی ایستگاه سینوپتیک ارومیه و نرم‌افزار REF-ET برای دوره ۲۰۰۱-۲۰۰۵ محاسبه نمودند. سپس با استفاده از داده‌های لایسیمتری به ارزیابی روش‌های فوق پرداخته و به این نتیجه رسیدند که روش بلانی کریدل نسبت به سایر روش‌ها مقدار تبخیر تعرق گیاه مرجع را با دقت بیشتری برآورد می‌نماید. Varkeshi et al. (2010) اقدام به شبیه‌سازی تبخیر و تعرق روزانه گیاه مرجع به روش هوش مصنوعی و روش‌های کلاسیک (تجربی) در مقایسه با اندازه‌گیری‌های لایسیمتری در اقلیم نیمه‌خشک سرد همدان نمودند. ایشان از دو روش هوش مصنوعی شبکه‌عصبی مصنوعی (ANN) و استنتاج عصبی- فازی تطبیقی (ANFIS) و دو روش

کلاسیک پنمن‌مانتیت- فائو ۵۶ (PMF56) و بلانی کریدل (BC) بهره جستند. در نهایت به این نتیجه رسیدند که روش هوش مصنوعی بهتر از روش‌های کلاسیک قادر به تخمین رفتار سیستم در همه مراحل عملکرد، چه مراحل افزایشی تبخیر و تعرق و چه مراحل کاهش تبخیر و تعرق می‌باشد. اما در مقایسه دو روش هوش مصنوعی نتایج شبکه عصبی- مصنوعی نسبت به مدل بهینه ANFIS بهتر می‌باشد. et al. (2014) Rezaei به ارزیابی کارایی مدل حداقل مربعات ماشین بردار پشتیبان در پیش‌بینی سطح ایستابی پرداختند. نتایج چهار ترکیب، نشان داد که عملکرد مدل LSSVM-RBF نسبت به مدل ANN برتری دارد. et al. (2015) Rezaei شبکه پایش سطح آب زیرزمینی را با استفاده از مدل حداقل مربعات ماشین بردار پشتیبان طراحی کردند. نتایج حاصل از این تحقیق نشان می‌دهد که ترکیب برتر مدل LS-SVM در برگیرنده شاخص‌های عملکرد است.

Yazdani and Ghahraman (2011) به تعیین بهترین روش تجربی برآورد تبخیر از سطح آزاد در اراضی شالیزاری آمل بر پایه آنالیز حساسیت و مقایسه آن با نتایج شبکه‌عصبی مصنوعی پرداختند. نتایج این تحقیق بیانگر کارایی بالای مدل پرسپترون چندلایه در برآورد تبخیر از سطح آزاد است. شایان ذکر است که مقادیر ضریب همبستگی، جذر میانگین مربع خطا و میانگین خطای مطلق در بهترین مدل به ترتیب برابر ۰/۹۵، ۰/۷۲ و ۰/۲۲ است. همچنین Irmak et al. (2002) در منطقه‌ای با اقلیم مرطوب، با به‌کارگیری روابط ضریب تست میزان تبخیر را به‌دست آوردند. (Snyder, 1992; Frevert et al. 1983) اقدام به برآورد ضریب تست و  $ET_0$  کردند. آن‌ها روشن ساختند که نتایج حاصل از رابطه (Frevert et al. 1983) به نتایج برآورد روش PMF-56 نزدیک‌تر است. در تحقیق دیگری Gundeekar et al. (2008) در کشور هند با استفاده از مدل‌های ضریب تست، مقدار  $ET_0$  را برآورد کردند. نتایج به‌دست‌آمده نشان داد که مدل اشنادر (SNYDER) بهترین مدل برای برآورد  $ET_0$  در مناطق مورد بررسی با اقلیم نیمه‌خشک هند است. Sabziparvar et al. (2010) به بررسی و ارزیابی مدل‌های تجربی ضریب تست برای برآورد  $ET_0$  در دو اقلیم خشک‌گرم و نیمه‌خشک‌سرد

در این تحقیق با استفاده از مدل LSSVM مقدار تبخیر دشت بیرجند برآورد شده و با تعدادی از روش‌های تجربی مقایسه شد.

## ۲- مواد و روش‌ها

### ۲-۱- موقعیت جغرافیایی دشت بیرجند

دشت بیرجند با مساحت ۳۹۴۹ کیلومترمربع در جنوب خراسان واقع است که در شمال و شمال‌غربی آن رشته‌کوه مؤمن‌آباد به ارتفاع ۲۷۸۰ متر و در جنوب آن رشته‌کوه باقران به ارتفاع ۲۷۲۰ متر از سطح دریا به‌گونه‌ای تقریباً موازی کشیده شده‌است. این شهر از شمال به شهرستان قائن و از جنوب به شهرستان نهبندان و کرمان از شرق به افغانستان و از غرب به شهرستان فردوس و طبس محدود می‌شود. دشت بیرجند دارای ۱۱ ایستگاه باران‌سنجی، یک ایستگاه اقلیم‌شناسی و یک ایستگاه سینوپتیک می‌باشد. دشت‌ها و زمین‌های هموار بیرجند در قسمت مرکزی و شمال شرق تا شهرستان و عمدتاً در فاصله سلسله کوه‌های باقران و مؤمن‌آباد که به‌صورت دو رشته موازی از شمال‌غربی به جنوب‌شرقی شهرستان امتداد یافته‌اند، قرار دارد. این دشت‌ها از نظر کشاورزی، گله‌داری و احداث راه‌های ارتباطی از اهمیت خاصی برخوردار هستند. از مهم‌ترین دشت‌ها، دشت بیرجند است که در فاصله دو کوه باقران و مؤمن‌آباد قرار دارد. برای انجام این تحقیق از آمار و اطلاعات هواشناسی سال‌های ۱۳۷۱ تا ۱۳۸۷ استفاده شده است. مقادیر روزانه داده‌ها با میانگین‌گیری از ۸ قرائت انجام و در طول شبانه‌روز محاسبه شد. این آمار و اطلاعات شامل دمای حداقل و حداکثر، رطوبت نسبی (کمینه و بیشینه)، ساعات آفتابی، میزان بارندگی و سرعت باد است (جدول ۱). مقادیر تبخیر از تشت به‌صورت روزانه برداشت و برای سنجش دقت سایر روش‌ها استفاده شد.

پرداختند. طی دهه اخیر مدل‌های شبکه‌عصبی مصنوعی (ANN, Artificial Neural Network) منطق فازی (Fuzzy Logic) و عصبی-فازی (Neuro-Fuzzy) به ابزاری برای پیش‌بینی تبدیل‌شده و در زمینه‌های گوناگون، نتایج بهتری نسبت به روش‌های قبلی به‌دست داده‌اند. (Chang et al. (2008) با استفاده از روش عصبی-فازی، تبخیر و تعرق گیاه مرجع را برآورد کردند. آن‌ها داده‌های تابش کل خورشیدی، سرعت‌باد، رطوبت نسبی و درجه حرارت را در حکم ورودی به کاربردند. نتایج این تحقیق بیانگر آن است که میزان خطای تبخیر-تعرق مرجع برآورد شده با این روش کمتر از روش PMF-56 است. در تحقیق دیگری که Landeras et al. (2008) در شمال اسپانیا به انجام رساندند، برآورد ETo با استفاده از شبکه‌عصبی-مصنوعی، روش‌های تجربی و نیمه تجربی مورد مقایسه قرار گرفت. نتایج آن‌ها نشان داد که روش شبکه‌عصبی بر دیگر روش‌ها برتری دارد. (Kumar et al. (2002) شبکه عصبی را ابزاری برتر و توانمندتر در مقایسه با روش‌های تجربی برای پیش‌بینی تبخیر و تعرق گیاه مرجع در دیویس کالیفرنیا دانستند. (Jensen et al. (1990) اهمیت ۲۰ روش مختلف را به‌منظور برآورد تبخیر و تعرق با داده‌های لایسیمتری برای ۱۱ ایستگاه در شرایط اقلیمی مختلف بررسی نمودند و روش پنمن مانیتث را بهترین روش برای تمامی وضعیت‌های اقلیمی معرفی نمودند. (Irmak et al. (2003) با استفاده از ایستگاه‌های هواشناسی در فلوریدا با اقلیم مرطوب، ۲۱ روش تخمین مقادیر روزانه ETo با مقادیر به‌دست‌آمده از روش پنمن-مانیتث - فائو مقایسه کردند. مدل‌های تابشی عملکرد بهتری از مدل‌های دمایی نشان دادند و از بین روش‌هایی که فقط از داده‌های دما استفاده می‌کردند، روش‌هاگریوز سامانی دارای کمترین مقدار خطا بود. بررسی‌ها نشان داد که مطالعات بسیار محدودی در ارتباط با توانایی مدل‌های ماشین‌بردار پشتیبان به‌عنوان یکی از روش‌های نوین در پیش‌بینی و شبیه‌سازی پدیده‌های طبیعی بالأخص تبخیر انجام شده است. لازم به ذکر است در این تحقیق قبل از بررسی کارایی مدل ماشین‌بردار پشتیبان بهترین ورودی مدل با آزمون گاما انجام شد که توانایی مدل ماشین‌بردار پشتیبان را افزایش خواهد داد.

جدول ۱- میانگین پارامترهای هواشناسی ایستگاه سینوپتیک بیرجند- خراسان جنوبی طی سال‌های ۱۹۹۲-۲۰۰۸

میانگین دما (°C)	حداقل دما (°C)	حداکثر دما (°C)	رطوبت نسبی نسبی کمینه (%)	رطوبت نسبی بیشینه (%)	میانگین رطوبت نسبی (%)	سرعت باد (m/s)	ساعات آفتابی (h)	میزان بارندگی (mm/day)
۱۶/۵۷	۸/۵۵	۲۴/۵۹۹	۱۹/۴۸	۵۲/۸۴	۳۶/۱۶	۴/۳۵	۹/۰۵	۰/۴۱۹

## ۲-۲-آزمون گاما

این آزمون نشان‌دهنده بخشی از واریانس خروجی است که نمی‌توان با اعمال هر مدل هموار بر ورودی‌ها محاسبه شود. برای درکی بهتر از این آزمون، چنانچه فرض شود دو نقطه  $x$  و  $x'$  در فضای ورودی متناظر با خروجی‌های  $y$  و  $y'$  به هم نزدیک شوند باید در فضای خروجی نیز این دو نقطه به هم نزدیک گردند در غیر این صورت، نزدیک نبودن این دو نقطه ناشی از اختلافی خواهد بود که توسط خطاها ایجاد می‌شود. محاسبه گاما شکل ساده‌ای از انحراف خطا به شمار می‌آید که مقدار خطای برآورد شده (واریانس خطا) را نسبت به داده‌های مستقیم نشان می‌دهد. مجموعه‌ای از داده‌های ورودی  $X_i$  و خروجی  $Y_i$  مشاهده شده از پدیده‌ها به صورت رابطه (۱) بیان می‌شود.

$$\{(x_i, y_i), 1 \leq i \leq M\} \quad (1)$$

$X \in R^M$  برای ورودی و  $Y \in R$  برای خروجی است.  $M$  تعداد سری مجموعه داده‌های مشاهده شده را نشان می‌دهد. پس آزمون گاما با فرض عدم قطعیت و پذیرش مقدار خطا در به‌دست‌آوردن خروجی از داده‌های ورودی به دلیل پیچیدگی و غیرخطی بودن پدیده‌های مدل‌سازی، این خطا به صورت رابطه (۲) نشان داده می‌شود:

$$Y = F(X_1, X_2, \dots, X_m) + \Gamma \quad (2)$$

$F$  نشان‌دهنده تابع همواری است که برای مدل‌سازی داده‌ها استفاده می‌گردد.  $\Gamma$  نشان‌دهنده متغیر تصادفی نمایش خطا است. این خطا ممکن است به دلایل عدم دقت اندازه‌گیری‌ها و هموار نبودن ارتباط اساسی بین ورودی و خروجی به وجود آید (Varkeshi et al. 2010). اگر میزان خطا زیاد باشد باید از هرگونه تلاش در جهت ایجاد مدلی هموار جلوگیری نمود. فرض می‌گردد میانگین خطا ( $\Gamma$ ) صفر است. علیرغم نامعلوم

بودن  $f$ ، مشروط به شرایط مطمئن، آزمون گاما تخمینی از واریانس خطا ( $r$ ) می‌باشد.  $\Gamma$  (مقدار آماره گاما) بر اساس  $k$ مین همسایه‌های نزدیک  $X_{N\{I,k\}}$  برای هر بردار ورودی  $X_i$  است. آماره گاما از تابع دلتا مخصوص بردارهای ورودی به صورت رابطه (۳) به دست می‌آید:

$$\Sigma_m = \frac{1}{m} \sum_{i=1}^m |X_{N\{i,k\}} - x_i|^2 \quad 1 \leq K \leq p \quad (3)$$

تابع مربوطه گاما برای هر خروجی  $y_i$  متناظر با بردار ورودی در رابطه (۱) بر اساس معادله (۴) به دست می‌آید:

$$Y_M = \frac{1}{2M} \sum_{i=1}^M |Y_{N\{i,k\}} - y_i|^2 \quad 1 \leq K \leq p \quad (4)$$

که مقدار  $y$  معین برای  $K$  امین همسایه نزدیک بردار ورودی  $X_i$  در رابطه (۱) است. میزان  $p$  را می‌توان بین ۵۰-۱۰ انتخاب کرد (Jones et al. 2002). با ایجاد رابطه رگرسیونی خطی رابطه (۵) به دست می‌آید:

$$Y = A\sigma + \Gamma \quad (5)$$

وقتی  $\sigma$  به سمت صفر میل نماید مقدار آماره گاما به صورت رابطه (۶) نشان داده می‌شود:

$$Y_M \rightarrow VAR(r)\sigma_M \rightarrow 0 \quad (6)$$

محاسبه شیب نیز به عنوان یک پارامتر بدون بعد هم می‌تواند اطلاعات مفیدی در رابطه با پیچیدگی مدل ارائه نماید. نسبت  $v$  نیز بررسی شکلی را که وابسته به دامنه خروجی است می‌سازد و خروجی مناسب برای توابع ساده را مدل‌سازی می‌کند که به صورت رابطه (۷) محاسبه می‌گردد (AbediKoupai et al. 2000).

$$S = \frac{\Gamma}{\sigma^2(Y)} \quad (7)$$

$\sigma^2(Y)$  واریانس خروجی را نشان می‌دهد. این نسبت بین ۱-۰ تغییر می‌کند که هرچه میزان آن به صفر نزدیک‌تر باشد بیانگر برآوردی مناسب‌تر از خروجی  $Y$  است. با استفاده از آزمون  $M$  می‌توان تعداد داده‌های موردنیاز را برای بخش

شبکه می‌گردد. به منظور افزایش کارایی مدل (LS-SVM)، همه داده‌ها ابتدا به شکل نرمال بین دو عدد ۰/۱ و ۰/۹ استاندارد شدند سپس آن‌ها بعد از استفاده در شبیه‌سازی به مقادیر اولیه برگشتند. (Dehghanisani et al. 2004)

$$X_i = 0/8 \left( \frac{x - x_{min}}{x_{max} - x_{min}} \right) + 0/1 \quad (11)$$

که در آن  $X_i$  مقدار استاندارد شده،  $x$  مقدار واقعی  $X_{min}$  و  $X_{max}$  به ترتیب مقادیر حداقل و حداکثر داده‌ها می‌باشند. در بخش مدل‌سازی، ۷۰ و ۳۰ درصد داده‌ها به ترتیب به داده‌های آموزشی و داده‌های بخش آزمون تعلق گرفت. در بخش اعتبارسنجی از تکنیک اعتبارسنجی تقاطعی با روش بهینه‌سازی سیمپلکس استفاده گردید. بعد از تعیین مقادیر ترکیب بهینه  $\gamma$  و  $\sigma$  با توجه به شکل (۱) مدل LSSVM ساخته شد. با استفاده از شرایط کان-تاگر (KKT) مدل حداقل مربعات ماشین‌بردار پشتیبان به منظور تابع تخمین به صورت رابطه (۱۲) نتیجه می‌شود:

$$Y(x) = \sum_{i=1}^n \alpha_i k(x_i, x_j) + b \quad (12)$$

$K(X_i, X_j)$  تابع کرنل نامیده می‌شود (رابطه ۱۳) که با تبعیت از شرایط مرسر ۱ تابعی با ایجاد ضرب داخلی در فضای ویژگی معرفی می‌شود (Shiloh shah 2007).

$$K(X_i, X_j) = \langle \varphi(X_i), \varphi(X_j) \rangle, \quad i, j = 1, \dots, N \quad (13)$$

آموزش تعیین نمود. در این تحقیق از نرم‌افزار WINGAMA استفاده شد.

### ۲-۳- مدل حداقل مربعات ماشین‌بردار پشتیبان

روش ماشین‌بردار پشتیبان بر پایه تئوری یادگیری آماری ارائه شد که از حداقل‌سازی ریسک ساختاری پیروی می‌کند. این مدل برخلاف مدل SVM به جای مسئله برنامه‌نویسی درجه دوم در حل از معادلات خطی استفاده می‌نماید، بنابراین دقت محاسباتی بالاتری نسبت به ماشین‌بردار پشتیبان کلاسیک دارد (Chang et al. 2008). مدل رگرسیونی LS-SVM از رابطه (۸) پیروی می‌کند:

$$y(x_i) = w^T \cdot Q(x_i) + b \quad (8)$$

که در آن  $w$  و  $b$  به ترتیب مقادیر وزن‌ها و اربیی تابع رگرسیون هستند که از طریق حداقل‌سازی تابع هدف در رابطه (۹) تعیین می‌گردند:

$$\min_{w, b} J(w, e) = \frac{1}{2} w^T w + \frac{\gamma}{2} \sum_{i=1}^n e_i^2 \quad (9)$$

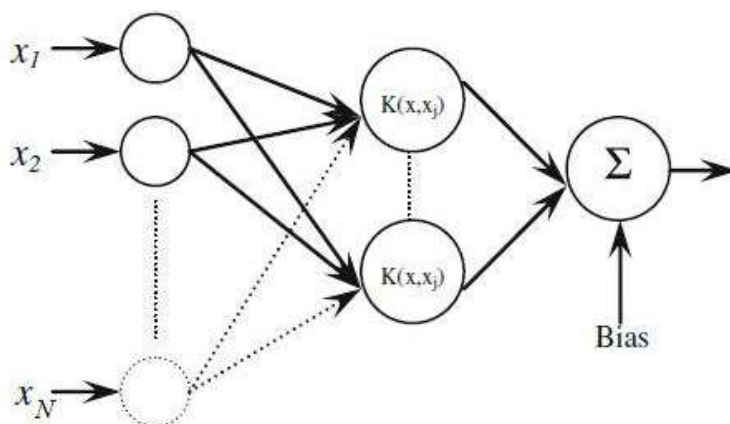
با محدودیت:

$$y_i = w^T \sigma(x_i) + b + x_i \quad (10)$$

که در آن‌ها  $\gamma$ : پارامتر تنظیم‌کننده بخش خطا است و  $e_i$ : خطای داده‌های آموزشی می‌باشد.

### ۲-۴- تجزیه و تحلیل داده‌ها

وارد کردن داده‌ها به صورت خام باعث کاهش سرعت و دقت



شکل ۱- معماری مدل LS-SVM (Mellit 2013)

این تحقیق از سه تابع کرنل مختلف که در جدول (۲) نشان داده شده، استفاده شد. اجزای مدل‌های ورودی در مدل LSSVM در جدول (۳) ذکر گردیده است.

برای حل، تعیین پارامتر تنظیم‌کننده  $\phi$  و تعیین  $\sigma < 0$  پارامتر مربوط به تابع کرنل (مربوط به کرنل RBF) نیاز است. انتخاب بهترین تابع کرنل به وسیله آزمون و خطا امکان‌پذیر است. در

جدول ۲- نام و فرمول توابع کرنل مورد استفاده در مدل LSSVM

فرمول توابع	نام توابع
$K(x_i, x_j) = x_i^t x_j$	Linear
$K(x_i, x_j) = (x_i^t x_j + \tau)^d$	Polynomial
$K(x_i, x_j) = \exp(- x_i - x_j ^2 / 2\sigma^2)$	Radial basis function

جدول ۳- اجزای مدل‌های ورودی مختلف در مدل LSSVM

نوع ترکیب	نام ترکیب
Tmin, Tmean, U@2, S	A
Tmin, U@2, Rmin, S	B
Tmean, U@2, Rmin, max, S	C
Tmin, U@2, Rmax, S	D
Tmean	E
Tmean, Rmean	F
Tmean, Tmax, Rmean, Rmax, U@2	G

$$E_s(tx) = 0.6018 \exp[17.27 * tx/tx + 237.3] \quad (17)$$

$$E_s(tn) = 0.6018 \exp[17.27 * tn/tn + 237.3] \quad (18)$$

$$E_a = 0.6018 \exp\left[\frac{REx + RHn}{2} / \frac{50}{es(tx)} + \frac{50}{es(tn)}\right] \quad (19)$$

$$E_a = 0.6018 \exp\left[\frac{REx + RHn}{2} / \frac{50}{es(tx)} + \frac{50}{es(tn)}\right] \quad (20)$$

در معادلات فوق  $e_s(tx)$  و  $e_s(tn)$  به ترتیب مقدار فشار بخار اشباع در دمای کمینه و بیشینه است.  $e_a$  نیز معرف مقدار فشار بخار واقعی هواست. (Alizadeh, 2000).

## ۲-۵- معیارهای ارزیابی و سنجش مدل

برای ارزیابی و سنجش مدل از ضریب تبیین  $R^2$ ، معیار ناش-ساتکلیف، (NSE) استفاده شد که روابط آن‌ها در زیر ارائه گردیده است.

$$R^2 = \left( \frac{\sum_{i=1}^n (Qoi - Qo)(Qsi - Qs)}{\sqrt{\sum_{i=1}^n (Qoi - Qo)^2} \sqrt{\sum_{i=1}^n (Qsi - Qs)^2}} \right)^2 \quad (21)$$

در این تحقیق از سه روش تجربی تخمین تبخیر به منظور ارزیابی نتایج تست تبخیر و ارزیابی مدل LSSVM استفاده شد. معادلات تجربی مورد استفاده در این تحقیق به صورت روابط (۱۴) و (۱۵) و (۱۶) است. (Yazdani and Ghahraman, 2011).

$$E = 0.0018 * (T + 25)^2 (100 - RH) \quad (14)$$

$$E = 0.833 * (4.57T + 43.3) \quad (15)$$

که  $E$  مقدار تبخیر از سطح آزاد آب برحسب میلی‌متر در ماه،  $T$  میانگین دما برحسب درجه سانتی‌گراد،  $RH$  میانگین رطوبت نسبی برحسب (درصد) می‌باشد.

$$E = 0.028 * U_2 * (e_s - e_a) \quad (16)$$

که  $U_2$  سرعت باد در ارتفاع دو متری برحسب (کیلومتر در ساعت)،  $e_s$  و  $e_a$  به ترتیب فشار بخار اشباع و فشار واقعی بخار آب در هوا برحسب (میلی‌متر جیوه) می‌باشد. به دلیل اینکه مقادیر فشار بخار اشباع و فشار واقعی هوا در این ایستگاه محاسبه نشده است، لذا مقادیر  $e_s$  و  $e_a$  از روابط زیر به دست می‌آید:

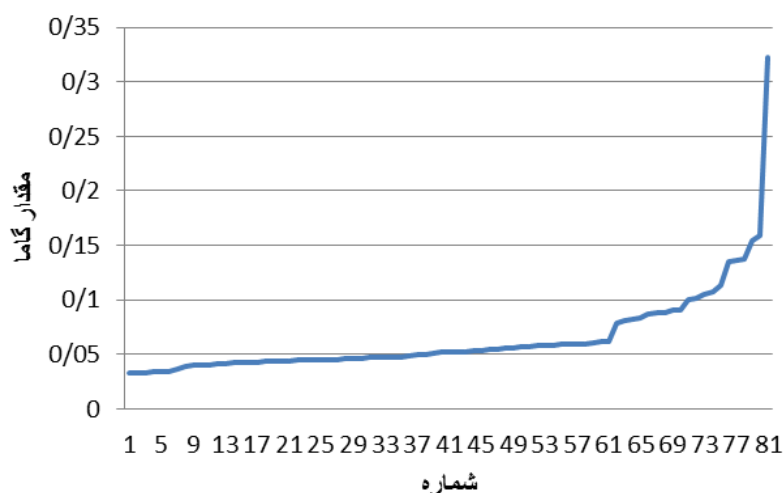
### ۳- یافته‌ها و بحث

#### ۳-۱- آموزش و اعتبار سنجی

در این مطالعه آثار حذف هر یک از پارامترها با استفاده از نرم‌افزار GAMATEST بر تقریب تبخیر بررسی شد.

#### ۳-۲- مرحله طراحی

در این تحقیق با استفاده از نرم‌افزار GAMATEST، ۹۰ ترکیب مختلف از پارامترهای ورودی مورد آزمون قرار گرفت. در نمودار شکل (۲) تغییرات گاما برای ۹۰ ترکیب نشان داده شده است.



شکل ۲- نمودار مقادیر گاما به ازای شماره داده

مقدار کمتر باشد مدل از پیچیدگی کمتری برخوردار است. خطای استاندارد، میزان اطمینان مدل را نشان می‌دهد. جدول (۵) نشان می‌دهد که از میان ترکیب‌هایی که به وسیله مدل LSSVM ارزیابی شده ترکیب‌های با تابع کرنل POLY با بالاترین میزان ضریب تبیین ( $R^2$ ) و کمترین میزان متوسط مطلق خطا و ریشه متوسط مربعات خطا بهترین ترکیب برای برآورد تبخیر است. همچنین ترکیب‌های با تابع کرنل LIN به‌عنوان بدترین ترکیب شناخته شد؛ چون دارای  $R^2$  کمتر و RMSE و MAE بالا بود. از بین مدل‌های تجربی روش ایوانف (رابطه ۱۴) بهترین شرایط را برای برآورد تبخیر و روش هنفر (رابطه ۱۶) کمترین شرایط را دارا بود.

$$NSE = 1 - \frac{\sum_{i=1}^n (Q_{si} - Q_{oi})^2}{\sum_{i=1}^n (Q_{oi} - Q_o)^2} \quad (22)$$

که در آن،  $Q_{oi}$  مقدار تبخیر مشاهداتی،  $Q_{si}$  مقدار تبخیر شبیه‌سازی شده،  $Q_o$  میانگین مقدار تبخیر مشاهداتی،  $n$  تعداد داده‌های مشاهداتی،  $Cc$  ضریب همبستگی خطی بین  $Q_o$  و  $Q_s$ ،  $\alpha$  نسبت انحراف معیار  $Q_s$  بر انحراف معیار  $Q_o$ ،  $\beta$  نسبت میانگین  $Q_s$  بر میانگین  $Q_o$  می‌باشد. در این تحقیق از جعبه‌ابزار LS-SVM1.8 در محیط متلب استفاده شده است. الگوریتم LS-SVM در دو مرحله آموزش و اعتبار سنجی و طراحی به کار رفت.

تعداد داده‌های مورد استفاده در این تحقیق ۶۲۱۱ می‌باشد. در جدول (۴) شاخص‌های ارزیابی عملکرد هر یک از ترکیب‌های ورودی آمده است. از آنجاکه مقدار ضریب تبیین ( $R^2$ ) برای ارزیابی مدل اکتفا نمی‌کند، باید همه شاخص‌های آماری محاسبه و در کنار هم بررسی شوند. به‌منظور گزینش بهترین ترکیب در جدول (۴)، میزان آماره گاما، شیب، خطای استاندارد، برای ۴ ترکیب برتر آورده شده است. که بر اساس کمترین مقدار گاما بهترین ترکیب انتخاب شد. در این تحقیق تنها میزان آماره گاما به‌عنوان معیار انتخاب بهینه مورد ارزیابی قرار گرفت. مقادیر شیب، خطای استاندارد و نسبت  $\nu$  نیز اطلاعاتی در مورد میزان پیچیدگی مدل ارائه می‌نمایند. مقدار شیب، شیب خط رگرسیونی را نشان می‌دهد که هر چه این



جدول ۴- مقادیر محاسبه‌شده توسط آزمون گاما

خطای استاندارد	شیب	مقدار گاما	ترکیب	نام ترکیب
۰/۰۰۰۵۱	۰/۱۴۶۴۲	۰/۰۳۲۷۱۲	Tmin,Tmean,U@2,S	A
۰/۰۰۰۵۷	۰/۱۰۰۲۵	۰/۰۳۲۸۲	Tmin,U@2,Rmin,S	B
۰/۰۰۰۴۶	۰/۰۶۴۶۴	۰/۰۳۳۳۲	Tmean,U@2,Rmin,max,S	C
۰/۰۰۰۶۹	۰/۰۹۱۴۰	۰/۰۳۴۴۸۱	Tmin,U@2,Rmax,S	D

جدول ۵- ارزیابی مدل LSSVM و مدل‌های تجربی

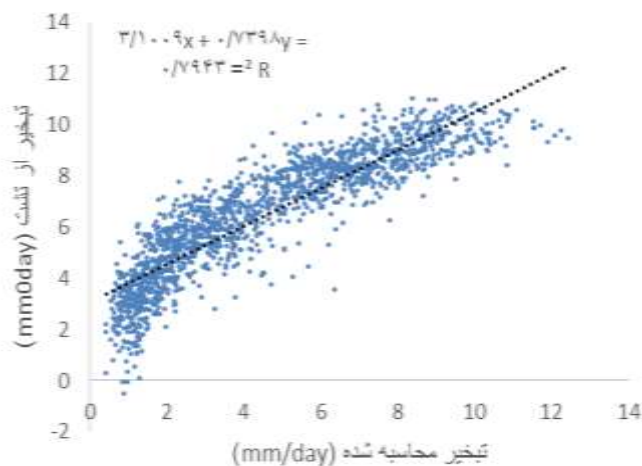
پارامتر $\gamma$	MAE	RMSE	$R^2$	نام ترکیب	مدل
۰/۰۲۸۹	۰/۹۱۹	۱/۱۷۰	۰/۸۵۳	A	LSSVM-LIN
۱۶/۱۰۹	۰/۷۵۲	۱/۰۰۵	۰/۸۸۹	A	LSSVM-RBF
۰/۰۰۰۰	۰/۷۴۹	۰/۹۹۵	۰/۸۹۱	A	LSSVM-POLY
۰/۲۹۴۳	۰/۹۳۳	۱/۱۹۰	۰/۸۵۳	B	LSSVM-LIN
۷/۴۷۱۸	۰/۷۶۳	۱/۰۲۳	۰/۸۹۴	B	LSSVM-RBF
۰/۰۰۰۱	۰/۷۶۴	۱/۰۱۷	۰/۸۹۴	B	LSSVM-POLY
۰/۳۶۱۵	۰/۹۲۳	۱/۱۷۲	۰/۸۴۸	C	LSSVM-LIN
۹/۶۰۰۲	۰/۷۵۶	۱/۰۰۶	۰/۸۹۱	C	LSSVM-RBF
۰/۰۰۵۸	۰/۷۵۵	۰/۹۹۸	۰/۸۹۲	C	LSSVM-POLY
۰/۳۹۲۶	۰/۹۲۲	۱/۱۷۶	۰/۸۵۷	D	LSSVM-LIN
۹/۶۴۷۹	۰/۷۵۷	۱/۰۱۰	۰/۸۹۲	D	LSSVM-RBF
۰/۰۰۰۱	۰/۷۵۸	۱/۰۱۰	۰/۸۹۱	D	LSSVM-POLY
۲/۳۹۷۱	۱/۰۷۲	۱/۳۹۶	۰/۸۰۱	E	LSSVM-LIN
۱/۶۳۹۶	۰/۹۵۶	۱/۲۸۲	۰/۸۵۱	E	LSSVM-RBF
۰/۰۰۰۲	۰/۹۵۳	۱/۲۷۸	۰/۸۵۲	E	LSSVM-POLY
۱/۰۹۱۴	۱/۰۶۰	۱/۳۸۳	۰/۸۰۵	F	LSSVM-LIN
۱۱/۶۸۷۹	۰/۹۲۹	۱/۲۵۱	۰/۸۵۵	F	LSSVM-RBF
۰/۰۰۱۶	۰/۹۴۱	۱/۲۶۵	۰/۸۵۳	F	LSSVM-POLY
۰/۰۷۲۴	۰/۹۸۱	۱/۲۴۳	۰/۸۲۸	G	LSSVM-LIN
۷/۴۸۱۱	۰/۸۰۸	۱/۰۶۵	۰/۸۷۵	G	LSSVM-RBF
۰/۰۰۰۲	۰/۸۱۵	۱/۰۷۸	۰/۸۷۲	G	LSSVM-POLY
-	۱/۱۲۵	۱/۱۳۴	۰/۸۲۶	H	روش ایوانف (رابطه ۱۴)
-	۱/۱۲۶	۰/۹۷۴	۰/۷۵۶	I	روش هنفر (رابطه ۱۶)
-	۲/۶۳۵	۲/۰۱۵	۰/۷۹۴	J	روش USBR (رابطه ۱۵)

بالاتر برابر ۰/۸۹۱ و میزان متوسط مطلق خطا (MAE) برابر ۰/۷۴۹ و ریشه متوسط مربعات خطا (RMSE) برابر ۰/۹۹۵ نسبت به تمام ترکیب‌های مدل‌سازی شده با روش‌های دیگر و دیگر توابع کرنل از دقت بالاتری برخوردار است. نکته قابل توجه این نتیجه این است که در ورودی A هم دمای

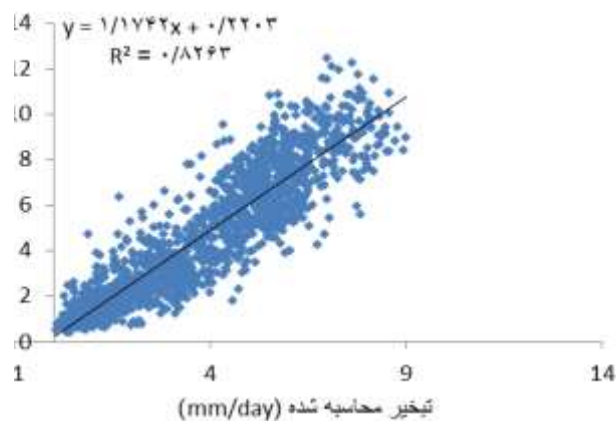
با توجه به جدول (۵) از بین ترکیب‌های بالا ترکیب‌های با تابع کرنل poly از دقت بالاتری برخوردار است. با توجه به شاخص‌های آماری ملاحظه می‌شود که ترکیب (A) با کمترین میزان آماره گاما، در مدل حداقل مربعات ماشین بردار پشتیبان (LSSVM) با تابع کرنل poly از میزان ضریب تبیین ( $R^2$ )

poly) نسبت به این مدل با تابع کرنل (Linear) و تابع کرنل (RBF) و مدل‌های تجربی مقادیر شبیه‌سازی شده بهتری را ارائه می‌دهد. این نتایج نشان می‌دهد تابع چندجمله‌ای poly نسبت به توابع دیگر می‌تواند ترکیب دقیق‌تری را ارائه نماید. شکل (۳) نتایج برآورد میزان تبخیر به روش‌های تجربی مورد مطالعه را نشان می‌دهد. در این شکل ضریب  $R^2$  بر روی هر کدام از نمودارها نشان داده شده است.

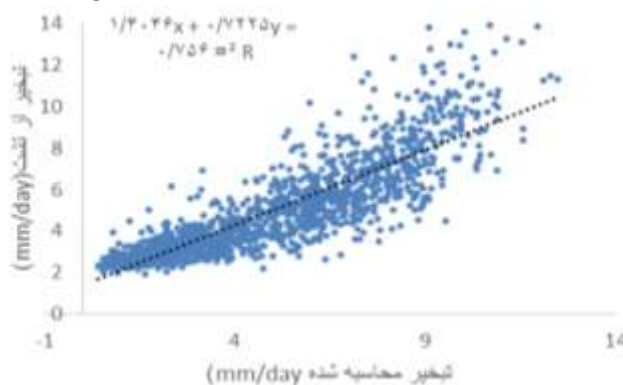
کمینه و هم‌دمای میانگین وجود دارد که در دیگر وردی‌ها این توأم وجود ندارد. با توجه به اهمیت دمای بیشینه و تأثیر آن بر دمای میانگین و برآورد تبخیر از اهمیت ویژه‌ای برخوردار است. از آنجاکه نوع تابع کرنل مناسب به صورت روش سعی و خطا تعیین می‌شود، اما در اکثر تحقیقات به این موضوع اهمیت داده نشده است. با توجه به جدول (۵) ملاحظه می‌گردد در ترکیب‌های A، B، C، D، E، F و G مدل حداقل مربعات ماشین بردار پشتیبان با تابع چندجمله‌ای



ج- محاسبه تبخیر به روش هنفر



الف- روش ایوانف



ب- روش USBR

شکل ۳- مقایسه دقت محاسبه تبخیر به روش‌های تجربی مورد مطالعه

انجام شده (Sabziparvar and Shademani (2012) با استفاده از روش‌های تجربی و شبکه عصبی مصنوعی و عصبی-فازی میزان تبخیر و تعرق گیاه مرجع را برای اقلیم‌های گرم و خشک برآورد کردند. نتایج حاصل از این تحقیق نشان داد

بررسی نتایج مدل‌های تجربی در شکل (۲) نشان داد که روش ایوانف (رابطه ۱۴) با دارا بودن میزان ضریب همبستگی بیشتر، از توانایی و دقت بیشتری در برآورد میزان تبخیر نسبت به دیگر روش‌ها برخوردار است. در تحقیقاتی که در این زمینه

که روش عصبی-فازی (ANFIS) نسبت به روش‌های دیگر نتایج بهتری در برآورد ضریب تشت و تبخیر و تعرق مرجع به‌دست می‌دهد. از بین روش‌های تجربی برآورد ضریب تشت روش کونینکا و اشنايدر بعد از روش‌های شبکه‌عصبی در اقلیم‌های گرم و خشک توصیه می‌شود. در تحقیق دیگری (Hozhbar et al. (2015) در ایستگاه کهرئز ارومیه میزان تبخیر و تعرق گیاه مرجع را با استفاده از مدل‌های تجربی و مدل‌سازی آن با شبکه‌عصبی مصنوعی به‌دست آوردند. نتایجی که از این تحقیق به‌دست آمد نشان می‌دهد که شبکه‌عصبی-مصنوعی نسبت به تمامی روش‌های کلاسیک عملکرد مناسب‌تری داشته‌است. همچنین در میان روش‌های کلاسیک مدل تورک در اولویت قرار دارد و روش‌های جنسن هیس، پنمن مانتیث و فائو ۵۶ در اولویت‌های بعدی قرار دارند. درحالی‌که در این تحقیق روش LSSVM برای دشت بیرجند نتایج بهتری ارائه می‌دهد.

#### ۴- نتیجه‌گیری

در این پژوهش کارایی مدل شبیه‌ساز LSSVM در پیش‌بینی تبخیر در مقایسه با روش‌های تجربی بررسی شد. نتایج زیر از این پژوهش قابل‌بیان است:

- ۱- مدل LSSVM با ترکیب پارامترهای دمای کمینه، دمای میانگین، سرعت باد و تعداد ساعات آفتابی در برآورد تبخیر دشت بیرجند در مقایسه با مدل‌های تجربی مناسب‌تر است.
- ۲- از میان توابع کرنل به‌کاررفته در مدل LSSVM تابع کرنل polynomial نتایج دقیق‌تری را نشان می‌دهد.
- ۳- مدل LSSVM مقادیر آماره‌های ضریب تبیین، جذر میانگین مربع خطا و میانگین خطای مطلق برای مجموعه داده‌ها به ترتیب ۰/۸۹۱، ۰/۷۴۹، ۰/۹۹۵ (میلی‌متر در روز) به‌دست آمد که در مقایسه با روش‌های دیگر ضریب همبستگی بزرگ‌تر و خطای کمتری داشت.
- ۴- مدل LSSVM با تابع کرنل polynomial بهترین روش در برآورد تبخیر می‌باشد.

## References

- AbediKoupai J., Islamian V. and Ameri M. (2000). Compare four methods of estimating evapotranspiration lysimeters in the area of micro-level reference data. Second National Conference on Manage of irrigation and drainage networks, Ahvaz. [in Persian]
- Adl F., Zeinalzadeh V. and Habebzadeh K.(2009). Estimating reference evapotranspiration using different methods (case study Urmia synoptic station). Second National Conference on Irrigation and Drainage network management, Ahvaz. [in Persian]
- Alizadeh A. (2000). Principles Applied Hydrology. Publication of Imam Reza.No:816. [in Persian]
- Chang F. J., Chang K. Y. and Chang L. C. (2008). Counter-propagation fuzzy-neural network for city flood control system. J. Hydrol., 358, 24-34.
- Dehghanisanij H. T., Yamamoto and V. Rasiah. (2004). Assessment of evapotranspiration estimation models for use in semiarid environments. Agri. Water Manag., 64, 91-106.
- Frevert, D. K., Hill, R. W. and Braaten, B. C. (1983). Estimation of FAO evapotranspiration coefficient. J. Irrig. Drain. Eng., 109(2), 265-270.
- Gundekar H. G., Khodke U. M. and Sarkar S. (2008). Evaluation of pan coefficient for reference crop evapotranspiration for semiarid region. J. Irrig. Sci., 26, 169-175.
- Hozhbar H., Moazed H. and Shokrikoochak S. (2015). Estimation of Reference evapotranspiration (ET<sub>o</sub>) using empirical models, artificial neural network modeling and their comparison with lysimeter data in Urmia Kahrizi station. J. Irrig. Water, 4(15), 13-25 [In Persian].
- Heidary M., Chalak A. and Khasheisuoki A. (2013). Experimental methods to determine the evaporation in semi-arid climates and mountainous Shiraz Bojnoord. National Conference and irrigation and evaporation of Kerman [in Persian].
- Irmak S., Haman D. and Jones J. W. (2002). Evaluation of class A pan coefficients for estimating reference evapotranspiration in humid location. J. Irrig. Drain. Eng., 128(3), 153-159.
- Irmak S. A., Irmak R. G., Allen and Jones J. W. (2003). Solar and net radiation based equations to estimate reference evapotranspiration in humid climates. J. Irrig. Drain. Eng., 129(5), 336-347.
- Jahanbakhshasl S., Movaheddanesh A. and Molavy V. (2001). Analysis models for estimating evapotranspiration stations in Tabriz. J. Agric. Know., 51-65. [In Persian]
- Jensen M. E., Burtman R. D. and Allen R. G. (1990). Evapotranspiration and irrigation water requirements, ASCE Manuals and Reports on Engineering Practices, No. 70, ASCE, New York.
- Kumar M., Raghuwanshi N. S., Singh R., Wallender W.W. and Pruitt W. O. (2002). Estimating evapotranspiration using artificial neural network. J. Irrig. Drain. ASCE, 128(4), 224-233.
- Landeras G., Ortiz-Barredo A. and Lopez J. J. (2008). Comparison of artificial neural network models and empirical and semi-empirical equations for daily reference evapotranspiration estimation in the Basque Country (Northern Spain). J. Agric. Water Manage., 95, 553-565.

- Rezaei A., Khasheisuoki A. and Shahedi A. (2015). Groundwater level monitoring network design using the least squares support vector machine (LSSVM). *Iran Soil Water Res.*, 45(4), 389-396. [In Persian]
- Rezaei A., Shahedi A., Khasheisuoki A. and RyahyMdvar H. (2014). Performance evaluation least squares support vector machine model to predict water table. *J. Irrig. Drain.*, 4(7), 510-520. [In Persian]
- Sabziparvar A. A. and Shadmani M. (2012). Evaluation of pan coefficients from ANN, ANFIS, and empirical methods, for estimation of daily reference. *J. Earth Space Phys.*, 38(1), 229-240. [In Persian]
- Sabziparvar A. A., Tabari H., Aeini A. and Ghafouri M. (2010). Evaluation of class A pan coefficient models for estimation of reference crop evapotranspiration in cold semi-arid and warm arid climates. *Water Resour. Manage.*, 24(5), 909-920.
- Shiloh Shah, R. (2007). Support vector machines for classification and regression. MSc Dissertation, Computer Science, McGill University Montreal , Quebec
- Snyder R. L. (1992). Equation for evaporation pan to evapotranspiration conversion, *J. Irrig. Drain. Eng.*, 118(6), 977-980.
- Varkeshi B., ZareAbyaneh H., Marufi A., Sabziparvar F. and Soltani M. (2010). Simulation of reference evapotranspiration using artificial neural method and empirical methods and comparison with experimental Lysimeter data in cold semi-arid climate of Hamedan. *J. Soil Water Conserv.*, 16(4), 79-100.
- Yazdani V. and Ghahraman B. (2011). Determine the best experimental methods for estimating evaporation from the free surface in paddy fields Amol based on sensitivity analysis and comparison with results of artificial neural network. *J. Water Res.*, 4(7), 47-58. [In Persian]

## Evaluation of LSSVM Model to Estimation of Evaporation and Its Comparison with Empirical Models

Mohadeseh Kavooosi<sup>1</sup>, Abbas Khasheisiuki<sup>2\*</sup>, Mohsen Pourrezabilondi<sup>3</sup>  
and Mohammad Hassan Najafi<sup>3</sup>

<sup>1</sup>M.Sc., Department of Water Engineering, Faculty of Agriculture, Birjand University, Birjand, Iran

<sup>2</sup>Associate Prof., Department of Water Engineering, Faculty of Agriculture, Birjand University, Birjand, Iran

<sup>3</sup>Assistant Prof., Department of Water Engineering, Faculty of Agriculture, Birjand University, Birjand, Iran

\*Corresponding author: abbaskhashei@birjad.ac.ir

### Original Paper

Received: May 30, 2017

Accepted: September 11, 2017

### Abstract

In this research, we used the meteorological parameters of Birjand Plain in South Khorasan Province during the 16-year period in order to evaluate the gamma test and to compare the accuracy of the model of least squares of the machine and experimental models to estimate evaporation. Using the gamma test method, among the parameters affecting evaporation, the optimal input parameters were determined for modeling the evapotranspiration from among 90 specified components. Seven superior components were obtained from other combinations. Then, the optimal combination was then evaluated using the least squares model of backup vector with various kernel functions and empirical methods. The results showed that the performance of the LSSVM-poly (polynomial) model, in combination with minimum temperature, average temperature, wind speed and sunny hours with performance indices ( $R^2 = 0.9995$  and  $RMSE = 995$ ) has a higher accuracy compared to the kernel functions and other experimental methods.

**Keywords:** Gamma Test; Evaporation; LSSVM; Birjand