نشریه آب و خاک
جلد ۲۴، شماره ۳، مرداد - شهریور ۱۳۸۹، ص.۶۲۴-۶۱۰

Journal of Water and Soil
Vol. 24, No. 3, July-Aug 2010, p. 610-624

ارزیابی تأثیر پیش‌پردازش پارامترهای ورودی به شبکه عصبی مصنوعی (ANNs) با استفاده از روش‌های رگرسیون گام به‌گام و گاما تست به منظور تخمین سریع‌تر تبخیر و تعرق روزانه

محمد قبائی سوق[۱] - ابوالفضل مساعدی*[۲] - موسی حسام[۳] - ابوطالب هزارجریبی[۴]

تاریخ دریافت:۸۸/۱۲/۲
تاریخ پذیرش:۸۹/۳/۳۰

چکیده

تبخیر و تعرق (ET_o) یکی از اجزای اصلی چرخه‌ی هیدرولوژیکی است که تعیین صحیح آن در مطالعات بیلان آبی، طراحی سیستم‌های آبیاری و برنامه‌ریزی و مدیریت منابع آب برای دست‌یابی به توسعه‌ی پایدار نقش به‌سزایی دارد. تبخیر و تعرق به علت نیاز به فاکتورهای اقلیمی مختلف و اثر متقابل این فاکتورها بر هم‌دیگر یک پدیده‌ی غیرخطی و پیچیده است. یکی از مراحل پیچیده در مدل‌سازی سیستم‌های غیرخطی، پیش‌پردازش پارامترهای ورودی جهت انتخاب ترکیبی مناسب از آن‌ها است. پیش‌پردازش داده‌ها سبب کاهش مراحل سعی و خطا و شناخت مهم‌ترین پارامترهای مؤثر بر پدیده‌ی مورد نظر جهت مدل‌سازی با استفاده از روش‌های هوشمند می‌گردد. در این تحقیق از دو روش رگرسیون گام‌به‌گام (FS) و گاماتست (GT) برای پیش‌پردازش پارامترهای ورودی به شبکه‌ی عصبی پرسپترون چندلایه (MLP) جهت تخمین ET_o روزانه‌ی ایستگاه سینوپتیک شیراز استفاده شده است. برای ارزیابی تأثیر پیش‌پردازش پارامترهای ورودی با استفاده از معیارهای مختلف آماری سنجش خطا به مقایسه‌ی دو مدل ANN-FS و ANN-GT (با پارامترهای پردازش شده) با یکدیگر و هم‌چنین با مدل ANN که هیچ‌گونه پیش‌پردازشی روی پارامترهای ورودی آن انجام نشده است، پرداخته شد. نتایج نشان داد که هر سه مدل از دقت بالایی جهت تخمین ET_o روزانه برخوردارند و از میان سه مدل فوق، مدل ANN-GT با مقدار ضریب تبیین (R^2) ۰/۹۹۹۵ و ریشه میانگین مربعات خطا (RMSE) ۰/۰۴۸۳ نسبت به دو مدل دیگر ANN-FS و ANN به ترتیب با مقادیر R^2 برابر با ۰/۹۹۸۴ و ۰/۹۹۹۴ و RMSE برابر با ۰/۰۸۷۴ و ۰/۰۵۴۸ از دقت بالاتری برخوردار است. اگرچه در این تحقیق برتری دقت مدل ANN-GT نسبت به مدل ANN ناچیز است اما توانایی‌های مشخص نمودن ترتیب اهمیت پارامترهای ورودی، تعیین تعداد تقریباً ۸۰۰ داده‌ی معنی‌دار برای آموزش شبکه و یافتن بهترین ترکیب که شامل همه‌ی پارامترهای ورودی به جز دمای حداکثر می‌باشد این آزمون را می‌تواند به عنوان ابزار مفیدی برای پیش‌پردازش پارامترهای ورودی جهت مدل‌سازی سریع‌تر تبخیر و تعرق تبدیل کند.

واژه های کلیدی : تبخیر و تعرق پتانسیل، شبکه‌های عصبی مصنوعی، گاماتست، رگرسیون گام‌به‌گام، ایستگاه سینوپتیک شیراز

مقدمه

تبخیر و تعرق یکی از اجزای اصلی چرخه‌ی هیدرولوژی است که تعیین صحیح آن در علوم آب از قبیل مطالعات توازن هیدرولوژیکی و طراحی و مدیریت سیستم‌های آبیاری از اهمیت بالایی برخوردار است (۱۲). علاوه بر این مقدار تبخیر و تعرق یکی از داده‌های اساسی در طراحی شبکه‌های آبیاری و زهکشی می‌باشد. زیرا طراحی شبکه‌های انتقال اعم از کانال‌های آبرسانی یا زهکشی و نیز سایر قسمت‌های طرح‌های آبی، وابسته به مقدار آب مورد نیاز از طریق پدیده‌ی تبخیر و تعرق است. هم‌چنین طراحی سیستم‌های آبیاری و تعیین بهترین دور آبیاری به تخمین مقدار آب مصرفی گیاه که شامل میزان تبخیر از سطح خاک و تعرق از گیاه است، نیازمند می‌باشد. لذا مساله‌ی تبخیر و تعرق باید به عنوان یکی از عوامل مهم و مؤثر در طراحی شبکه‌های آبیاری و زهکشی مدنظر قرارگیرد (۲). تخمین بیش از حد آب مورد نیاز گیاه ضمن هدر دادن آب آبیاری باعث ماندابی شدن اراضی، شستشوی مواد غذایی خاک و آلوده نمودن منابع آب زیرزمینی می‌شود. ضمن آن‌که تخمین کمتر نیز باعث اعمال استرس رطوبتی به گیاه شده و در نتیجه کاهش محصول را به همراه خواهد داشت (۹).

انتقال آب بصورت بخار از سطح خاک را تبخیر[5] و از سطح

۱، ۲، ۳ و ۴ – دانشجوی کارشناسی ارشد، دانشیار و استادیاران گروه مهندسی آب، دانشگاه علوم کشاورزی و منابع طبیعی گرگان
(* - نویسنده مسئول: Email: mosaedi@yahoo.com)

5- Evaporation

گیاهان را تعرق[1] گویند. از نظر فیزیکی این دو فرآیند مشابه هستند، زیرا در هر دو صورت تغییر حالت آب از مایع به بخار و انتقال آن به جو رخ می‌دهد. مجموع این دو تلفات را تبخیر وتعرق[2] گویند (۶). بر اساس استاندارد فائو، تبخیر و تعرق گیاه مرجع عبارت است از میزان آبی که یک مزرعه پوشیده از گیاه مرجع (نظیر چمن) در یک دوره زمانی مشخص مصرف نماید به طوری‌که گیاهان این مزرعه در طول دوره رشد با کمبود آب مواجه نشوند (۱۰).

اثر متقابل فاکتورهای کلیماتولوژی از قبیل دما، رطوبت نسبی، سرعت باد، تشعشع، نوع و مرحله رشد گیاه و عوامل دیگر سبب شده است تا تبخیر و تعرق به‌صورت پدیده‌ای غیرخطی و پیچیده بیان شود (۹). به‌خاطر دشواری و پیچیدگی فرآیند مذکور از سوی محققان معادلات تجربی مختلفی شامل: پنمن ۱۹۴۸[3]، تورنت‌وایت ۱۹۴۸[4]، بلانی کریدل ۱۹۵۰[5]، تورک ۱۹۶۱، هارگریوز-سامانی ۱۹۸۵[6]، جنسن هیز ۱۹۶۳[7]، پنمن- مانتیث ۱۹۶۵[8] و فائو پنمن- مانتیث ۱۹۹۸[9] ارائه شده است (۶).

پیچیدگی فرآیند تبخیر و تعرق و کثرت اطلاعات مورد نیاز برای محاسبه‌ی آن از یک سو و مشکلات موجود بر سر راه اندازه‌گیری این داده‌ها که غالباً موجب فقدان آن‌ها می‌گردد از طرف دیگر، ضرورت استفاده از شبکه‌های عصبی مصنوعی را در برآورد تبخیر و تعرق آشکار می‌سازد (۹). در طی سالیان اخیر تحقیقات متعددی در ارتباط با تخمین تبخیر وتعرق از روی داده‌های هواشناسی با استفاده از روش‌های هوشمند در نقاط مختلف دنیا انجام گرفته است. ادهیامبو و همکاران (۲۰۰۱)، کومار و همکاران (۲۰۰۲)، سیلوا (۲۰۰۲) و تراجکویچ و همکاران (۲۰۰۳) از داده‌های هواشناسی یکسانی در تحقیقات خود استفاده نمودند و طی تحقیقات خود نشان دادند که نتایج حاصل از تخمین تبخیر و تعرق با استفاده از شبکه‌های عصبی مصنوعی دارای دقت بالاتری نسبت به روش‌های رایج تبخیر و تعرق می‌باشند (۲۴). سادهر و همکاران (۲۶) برای محاسبه‌ی تبخیر و تعرق از داده‌های ساده شده مانند دمای هوا، تابش برون زمینی و ساعات آفتابی استفاده نموده و نتایج حاصل را رضایت‌بخش توصیف نمودند. کیسی و اوزتورک (۱۷) جهت تخمین مقدار تبخیر و تعرق به روش فائو پنمن-مانتیث از سیستم استنتاج عصبی فازی (ANFIS) استفاده نمودند.

شایان‌نژاد و همکاران (۸) با استفاده از اطلاعات ۵ ساله‌ی اقلیمی و لایسیمتری ایستگاه اکباتان همدان تبخیر و تعرق را به سه روش رگرسیون فازی، شبکه‌های عصبی مصنوعی و روش فائو پنمن-مانتیث محاسبه و با داده‌های حاصل از لایسیمتر مقایسه و بیان داشتند که روش رگرسیون فازی با ضریب تبیین ۰/۸۸ و مجذور میانگین مربعات خطایی برابر ۰/۷۴ میلیمتر بر روز نتایج بهتری را بدست می‌دهد. کوچک زاده و بهمنی (۹) در تحقیقی به ارزیابی عملکرد شبکه‌های عصبی مصنوعی در کاهش پارامترهای مورد نیاز برای برآورد تبخیر و تعرق مرجع در ایستگاه مهرآباد تهران و مقایسه‌ی آن با روش‌های فائو پنمن-مانتیث، بلانی کریدل، جنسن هیز اصلاح شده، و هارگریوز سامانی پرداختند. آن‌ها بدین منظور به کمک پارامترهای مختلف ۱۱ مدل مختلف ANN طراحی نمودند و با مرجع قرار دادن روش فائو پنمن-مانتیث از طریق معیارهای آماری خطا به ارزیابی مدل‌های ایجاد شده پرداختند. بر اساس نتایج بدست آمده دو پارامتر دما و سرعت باد به عنوان مؤثرترین فاکتورها بر دقت برآورد مدل‌های ANN شناسایی شد. زارع ابیانه و همکاران (۴) برای پیش‌بینی تبخیر و تعرق گیاه مرجع در منطقه همدان از شبکه عصبی مصنوعی و سیستم استنتاج عصبی فازی استفاده نمودند. آن‌ها در تحقیق خود با ایجاد ۴ مدل مختلف از ترکیب ۶ پارامتر دماهای حداکثر و حداقل، مقادیر رطوبت نسبی حداقل و حداکثر، سرعت باد و ساعات آفتابی بیان داشتند که از میان مدل‌های بررسی شده، مدل با ترکیب پارامترهای دمای حداقل، دمای حداکثر و ساعات آفتابی روزانه از دقت بالاتری جهت تخمین تبخیر و تعرق برخوردار است. هم‌چنین در تحقیقی دیگر زارع ابیانه و همکاران (۵) با استفاده از داده‌های لایسیمتری دقت شبکه عصبی MLP را جهت تخمین مقدار تبخیر و تعرق گیاه سیر با استفاده از پارامترهای هواشناسی، مورد ارزیابی قرار دادند. نتایج آن‌ها نشان داد که شبکه‌ی MLP از دقت خوبی برخوردار است و نسبت به پارامتر دمای حداکثر بیشترین حساسیت و حداقل رطوبت نسبی کم‌ترین حساسیت را داراست.

به‌طور کلی مدل‌سازی داده‌های خروجی با استفاده از مجموعه داده‌های ورودی همواره دارای مقداری خطا خواهد بود که منشأ ایجاد این خطا می‌تواند ناشی از: عدم دقت در اندازگیری مقادیر پارامترها، عدم کفایت دانش بشری برای اندازه‌گیری و تشخیص تمام عوامل مؤثر پدیده‌ی مورد نظر و عدم وجود رابطه‌ی هموار[10] بین مجموعه داده‌های ورودی و خروجی باشد (۱۵).

در شرایطی که هیچ دانشی از معادلاتی که رفتار سیستم را مشخص سازد در دسترس نباشد، گاماتست اساسا به عنوان یک ابزار، می‌تواند برای کمک به ساخت مدلی هموار از رفتار سیستمی آن پدیده بر اساس سری داده‌های اندازه‌گیری شده بکار رود. هم‌چنین این

1- Transpiration
2- Evapotranspiration
3- Penman
4- Thornth-Waite
5- Blaney-Criddle
6- Hargreavs-Samani
7- Jensen-Haise
8- Penman-Monteith
9- FAO Penman-Monteith
10- smooth

آزمون می‌تواند تا اندازه‌ای برای تخمین آن بخش از واریانس داده‌های خروجی که از روی داده‌های ورودی با ایجاد یک مدل هموار نمی‌توانند محاسبه شوند، به‌کار رود (۱۳).

این آزمون نخستین بار توسط کانسکر[1] ۱۹۹۷ و کمی بعد استفانسون[2] و همکاران در سال ۱۹۹۷ به صورت خلاصه معرفی شد و در سال‌های بعد توسط چوزانوا[3] و همکاران ۱۹۹۸، الیویرا[4] ۱۹۹۹، تسو[5] ۱۹۹۹، دورانت[6] ۲۰۰۱، تسو و همکاران ۲۰۰۲، جونس[7] و همکاران (۲۰۰۲) با جزئیات بیشتر مورد بررسی و استفاده قرار گرفت. ایوانس (۲۰۰۱) و ایوانس و جونس (۲۰۰۲) ضمن استفاده از این آزمون به اثبات آن از طریق روابط ریاضی پرداختند (۲۵). طی سالیان اخیر تحقیقاتی در ارتباط با استفاده از گاماتست بویژه کاربرد آن در تبخیر و تعرق منتشر شده‌است. ریمسان[8] و همکاران(۲۵) با استفاده از گاماتست عوامل مؤثر بر تابش خورشیدی را در حوضه آبریز برو[9] انگلستان تعیین نمودند. آن‌ها با استفاده از آمار پارامترهای روزانه هواشناسی شامل: دما، باران، سرعت باد و تابش برون زمینی و ایجاد مدل‌های مختلف از ترکیب پارامترهای فوق بهترین مدل را مدلی تشخیص دادند که دارای تمام پارامترهای فوق باشد و سپس با استفاده از دو مدل رگرسیون و شبکه‌ی عصبی به مدل‌سازی تابش خورشیدی پرداختند که نتایج حاکی از دقت بیشتر مدل رگرسیونی ایجاد شده بود.

مقدم‌نیا و همکاران (۱۹) برای مدل‌سازی تبخیر روزانه منطقه‌ی چاه‌نیمه، واقع در استان سیستان و بلوچستان از شبکه‌های عصبی مصنوعی و سیستم استنتاج عصبی فازی استفاده نمودند. آن‌ها در تحقیق خود از چهار پارامتر هواشناسی شامل: دمای هوا، کمبود فشار بخار اشباع، رطوبت نسبی و سرعت باد در ارتفاع دو متری استفاده نموده و با استفاده از گاماتست بهترین ترکیب که شامل پارامترهای کمبود فشار بخار اشباع، رطوبت نسبی و سرعت باد در ارتفاع دو متری می‌باشد را بدست آوردند. هم‌چنین با استفاده از آزمون M تست تعداد داده‌های مورد نیاز برای مدل‌سازی تبخیر را بدست آورند و در پایان بیان داشتند که نتایج حاصل از شبکه‌ی عصبی برای پیش‌بینی تبخیر نسبت به مدل‌های تجربی و سیستم استنتاج فازی عصبی از دقت بالاتری برخوردارند.

نوری و همکاران (۲۰) عوامل مؤثر بر میزان تولید زباله‌های هفتگی در شهر مشهد را با استفاده از دو تکنیک گاماتست و تحلیل مؤلفه‌های اصلی برآورد نمودند و سپس با استفاده از شبکه عصبی مصنوعی به مدل‌سازی این پدیده پرداختند. آن‌ها در این تحقیق با استفاده از دو روش فوق توانستند تعداد ۱۳ متغیر اولیه‌ی تاثیرگذار بر تولید زباله‌های هفتگی در مشهد را به ترتیب به ۵ و ۷ متغیر برای ورود به شبکه‌ی عصبی پرسپترون کاهش دهند. در پایان بیان داشتند که مدل حاصل از گاماتست و شبکه عصبی از دقت بیشتری برای مدل‌سازی تولید زباله‌های هفتگی برخوردار است. نوری و همکاران (۲۱) با استفاده از گاماتست غلظت مونواکسید کربن روزانه در ایستگاه قلهک تهران را به کمک شبکه‌ی عصبی مصنوعی و سیستم استنتاج فازی عصبی مدل‌سازی نمودند. آن‌ها با استفاده از دو تکنیک گاماتست و رگرسیون گام‌به‌گام توانستند تعداد ۱۲ متغیر اولیه‌ی تأثیرگذار بر غلظت مونواکسیدکربن روزانه را به ترتیب به ۹ و ۷ متغیر برای ورود به شبکه‌های عصبی مصنوعی و سیستم استنتاج فازی عصبی کاهش دهند. در پایان بیان داشتند که مدل حاصل از رگرسیون گام‌به‌گام و شبکه عصبی از دقت بیشتری برای مدل‌سازی غلظت مونواکسید کربن روزانه برخوردار است. احمدی و همکاران (۱۱) روش‌های متعدد گزینش پارامترهای ورودی مانند (گاماتست، نظریه‌ی آنتروپی، AIC[10] و BIC[11]) بر انتخاب پارامترهای مؤثر بر تابش خورشیدی در حوضه آبریز برو انگلستان را مورد بررسی قراردادند. آن‌ها ضمن مقایسه‌ی روش‌های مختلف بیان داشتند که قابلیت انتخاب بهترین ترکیب از ورودی‌ها، گاماتست را نسبت به سایر روش‌ها متمایز می‌سازد. هرچند این تکنیک تعداد داده‌های مورد نیاز برای مدل‌سازی را بطور قابل ملاحظه‌ای کم برآورد می‌نماید که در این زمینه نظریه‌ی آنتروپی دارای قابلیت بهتری است.

پیش‌پردازش پارامترهای ورودی جهت انتخاب ترکیب مناسب از میان آن‌ها یکی از فرایندهای پیچیده در مدل‌سازی سیستم‌های غیرخطی است. علیرغم وجود مقالات فراوان برای مدل‌سازی پدیده‌ی تبخیر و تعرق با استفاده از روش‌های هوشمند، هم‌چنان در این باره چند سؤال اساسی وجود دارد از جمله: کدام یک از داده‌های ورودی به مدل از اهمیت بیشتری برخوردار است؟ چه ترکیبی از داده‌های ورودی باید به مدل معرفی شود؟ چه تعداد داده برای مدل‌سازی این پدیده مورد نیاز است؟ بررسی منابع نشان می‌دهد که در اکثر مراجع برای مدل‌سازی تبخیر و تعرق با استفاده از روش‌های هوشمند، بدون بکارگیری روش‌های پیش‌پردازش پارامترهای ورودی، با استفاده از روش سعی و خطا، از میان ترکیبات محدود بررسی شده بهترین ترکیب انتخاب می‌شود. بنابراین فقدان یک روش قاعده‌مند برای یافتن بهترین ترکیب از میان پارامترهای مختلف ورودی احساس می

1- Koncar
2- Stefánsson
3- Chuzhanova
4- Oliveira
5- Tsui
6- Durrant
7- Jones
8- Remesan
9- Brue

10- Akaike's information criterion
11- Bayesian information criterion

شود که این تحقیق به دنبال معرفی روشی جهت تعیین بهترین ترکیب از میان پارامترهای مختلف ورودی در برآورد تبخیر و تعرق است. هم‌چنین اهداف دیگر این تحقیق ارزیابی تأثیر دو روش پیش‌پردازش پارامترهای ورودی (رگرسیون گام‌به‌گام و گاماتست) برای انتخاب بهترین ترکیب، مشخص نمودن ترتیب اهمیت پارامترهای ورودی مؤثر بر تبخیر و تعرق و تعیین تعداد داده‌های معنی‌دار برای ایجاد یک مدل هموار برای مدل‌سازی تبخیر و تعرق با استفاده از شبکه عصبی MLP می‌باشد.

مواد و روش‌ها

در این تحقیق برای انجام رگرسیون گام‌به‌گام از نرم‌افزار آماری R نسخه‌ی ۲/۹ استفاده شد. این نرم‌افزار یک نرم‌افزار رایگان با کتابخانه‌ی گسترده و قدرت گرافیکی بالا در مقایسه با دیگر نرم‌افزارهای آماری نظیر SAS، Minitab، SPSS است که توسط رابرت[1] و رز[2] از بنیان‌گزاران هسته‌ی اولیه‌ی R در دانشگاه نیوزلند برنامه‌نویسی شد (۲۷). هم‌چنین برای انجام گاماتست از نرم‌افزار وین‌گاما[3] نسخه‌ی ۱/۹۸ که توسط جمعی از محققان گروه کامپیوتر دانشگاه کاردیف[4] برای مدل‌سازی سیستم‌های غیرخطی طراحی گردیده‌است، استفاده شد (۱۴). برای مدل‌سازی شبکه‌های عصبی مصنوعی نیز از نرم‌افزار Matlab نسخه‌ی۷/۲ استفاده شده است. این نرم‌افزار امکان طراحی، ساخت، یادگیری و ارزیابی شبکه‌های عصبی مصنوعی را در اختیار می‌دهد و در برگیرنده‌ی شبکه‌های متفاوت با قوانین یادگیری مختلف می‌باشد.

موقعیت منطقه و داده‌های مورد استفاده

برای انجام این تحقیق از داده‌های ایستگاه سینوپتیک شیراز واقع در استان فارس با مختصات جغرافیایی طول ۵۲ درجه و ۳۲ دقیقه شرقی، عرض ۲۹ درجه و ۳۶ دقیقه شمالی، ارتفاع ۱۴۸۴ متر از سطح دریا و متوسط بارندگی سالانه ۳۴۶ میلی‌متر استفاده شده‌است. داده‌های مورد استفاده شامل: دمای حداکثر (T_{max})، دمای حداقل (T_{min})، رطوبت نسبی (RH)، ساعات آفتابی (R_s) و سرعت باد در ارتفاع دو متری (U_2) به صورت روزانه از ابتدای سال ۱۹۹۲ میلادی تا پایان سال ۲۰۰۷ میلادی می‌باشند که از سازمان هواشناسی کشور اخذ گردید. خصوصیات آماری داده‌های مورد استفاده در طول دوره آماری در جدول ۱ ارائه شده‌است.

1- Robert Gentlman
2- Ross Ihaka
3- Win Gamma
4- Cardiff

شبکه‌های عصبی مصنوعی

آغاز نظریه‌ی شبکه‌های عصبی مصنوعی تقریباً به دهه‌ی ۴۰ میلادی بر می‌گردد، زمانی که مک‌کلاچ[5] روانشناس معروف و والترپیتز[6] ریاضی‌دان در سال ۱۹۴۳ آن را پایه‌گذاری نمودند (۱۸). از دیدگاه ریاضی یک شبکه عصبی به عنوان یک تقریب‌گر است و توانایی آن در تقریب بین الگوهای یک مساله باعث می‌شود تا بتوان مسائل با پیچیدگی زیاد از قبیل شناسایی الگو[7]، تفکیک الگو[8]، نگاشت غیرخطی[9]، حافظهٔ انجمنی[10]، خودسازمان‌دهی[11] و کنترل را انجام دهد. اگر چه ایده‌ی شبکه عصبی مک‌کلاچ و پیتز بیشتر از ۶۰ سال پیش ارائه شد ولی نخستین کاربرد عملی شبکه‌های عصبی مصنوعی توسط روزنبلات[12] با معرفی شبکه‌های پرسپترون چند لایه[13] (MLP) انجام شد. بطور کلی ساختار شبکه‌ی عصبی MLP از سه لایه‌ی ورودی، پنهانی و خروجی تشکیل شده که برای معماری شبکه در هر لایه تعدادی نرون در نظر گرفته می‌شود. تعداد نرون‌های موجود در لایه‌های ورودی و خروجی با توجه به ماهیت مسأله‌ی مورد بررسی مشخص می‌شود، حال آن که تعداد نرون‌های موجود در لایه‌های پنهانی و هم‌چنین تعداد این لایه‌ها با سعی و خطا در جهت کاهش مقدار خطا توسط طراح مشخص می‌گردد (۱۹).

در این تحقیق برای مدل‌سازی ET_o با استفاده از شبکه‌ی MLP، از یک لایه‌ی پنهان با تعداد نرون‌های متفاوت و توابع محرک کران‌دار تابع سیگموید و از میان روش‌های مختلف آموزش به روش پس‌انتشار خطا با الگوریتم لونبرگ-مارکوارت، به دلیل همگرایی سریع‌تر در آموزش شبکه استفاده شده‌است. علت انتخاب یک لایه پنهان این است که چون لایه‌های میانی در ارتباط مستقیم با خروجی نیستند، بنابراین تغییرات این لایه‌ها تأثیر چندانی در تعدیل وزن‌ها ایجاد نمی‌کنند (۲۰). اساس روش پس‌انتشار خطا بر پایه قانون یادگیری اصلاح خطا می‌باشد که از دو مسیر اصلی رفت و برگشت تشکیل می‌شود. در مسیر رفت، بردار ورودی به شبکه اعمال شده و تأثیراتش از طریق لایه‌های میانی به لایه خروجی انتشار می‌یابد و بردار خروجی پاسخ واقعی شبکه را تولید می‌نماید.

5- Mcculloch
6- Walter Pitts
7- Pattern recognition
8- Pattern classification
9- Nonlinear mapping
10- Associative memory
11- Self organization
12- Rosenblatt
13- Multi layer perceptron

جدول ۱- برخی از پارامترهای آماری داده‌های ایستگاه سینوپتیک شیراز در طول دوره آماری ۲۰۰۷-۱۹۹۲

پارامتر / خصوصیت آماری	دمای حداکثر °C	دمای حداقل °C	دمای متوسط °C	رطوبت نسبی درصد	ساعات آفتابی ساعت	سرعت باد متر بر ثانیه	تابش برون زمینی MJ m^{-2} d^{-1}
حداکثر	۴۳/۲	۲۸/۶	۳۵/۶	۹۸/۵	۱۶/۲	۸/۳	۴۱/۱
حداقل	۱/۴	-۹/۶	-۳/۷	۱۱	۰	۰	۲۰/۱
میانگین	۲۶/۱	۱۱/۲	۱۸/۶	۳۹/۳	۹/۳	۲	۳۱/۷
انحراف معیار	۹/۶۰	۷/۸۰	۸/۵۴	۱۸/۵۲	۳/۱۱	۱/۱۶	۷/۵۱
ضریب تغیرات	۰/۳۶۸	۰/۶۹۸	۰/۴۵۹	۰/۴۷۲	۰/۳۳۵	۰/۵۶۹	۰/۲۳۷
چولگی	-۰/۱۸۷	-۰/۰۴	-۰/۰۸	۰/۹۴۹	-۱/۵۳	۰/۸۵۶	-۰/۲۱۱

معادله فائو پنمن-مانتیث (E_{PM56})

روش موسوم به فائو پنمن-مانتیث تکامل یافته‌ی روش ترکیبی پنمن-مانتیث ۱۹۶۵ است که توسط آلن و همکاران (۱۲) برای محاسبه‌ی تبخیر و تعرق در نشریه‌ی شماره ۵۶ سازمان فائو با نام E_{PM56} به صورت رابطه‌ی ۱ ارائه شده‌است. این روش با درجه اعتماد بالایی در دامنه وسیعی از مناطق و اقلیم‌ها برآورد صحیحی از تبخیر و تعرق گیاه مرجع ارائه می‌کند و از سوی سازمان خوار و بار جهانی (FAO) به عنوان تنها روش استاندارد برای محاسبه تبخیر و تعرق گیاه مرجع از روی داده‌های اقلیمی و هم‌چنین برای ارزیابی سایر روش‌ها پیشنهاد شده است (۱۶). با استفاده از این روش می‌توان ET_o را به صورت ماهانه، ۱۰روزه و روزانه محاسبه کرد.

$$E_{PM56}=\frac{0.408\,\Delta(R_n-G)+\gamma\left(890/(T_{mean}+273)\right)U_2(e_s-e_a)}{\Delta+\gamma(1+0.34\,U_2)} \quad (۱)$$

در رابطه‌ی۱ E_{PM56} تبخیر و تعرق گیاه مرجع (mm.day^{-1})، Δ شیب منحنی فشار بخار اشباع (kPa C^{-1})، R_n تابش خالص در سطح پوشش گیاهی (MJ m^{-2} d^{-1})، G چگالی شار گرمای خاک (MJ m^{-2} d^{-1})، γ ضریب سایکرومتری رطوبتی (kPa C^{-1})، T_{mean} میانگین دمای روزانه (°C)، U_2 سرعت باد در ارتفاع دو متری (m s^{-1})، e_s فشار بخار اشباع (kPa)، e_a فشار بخار واقعی (kPa) است.

رگرسیون گام‌به‌گام

زمانی که تعداد عوامل ورودی مؤثر (N) بر پدیده‌ای کم باشند، می‌توان با بررسی تمام ترکیبات ممکن و مقایسه‌ی خطا در مدل‌های مختلف رگرسیونی، بهترین مدل رگرسیونی برای پیش‌بینی آن پدیده را بدست آورد. اما وقتی که تعداد پارامترهای ورودی افزایش می‌یابد ارزیابی کلیه‌ی رگرسیون‌های ممکن به حجم محاسبات زیادی نیاز دارد. بنابراین روش‌های مختلفی ایجاد شده‌است که صرفاً تعداد کمی از مدل‌های رگرسیون دارای زیر مجموعه‌ای از متغیرها را در یک زمان بررسی می‌کنند. این روش‌ها عموماً به روش‌های نوع گام‌به‌گام[1] معروف هستند. روش‌های گام‌به‌گام را می‌توان در سه دسته کلی گزینش پیش‌رونده (FS)[2]، حذف پس‌رونده[3] و رگرسیون گام‌به‌گام که ترکیبی از روش‌های پیشین است تقسیم نمود (۳). محققان مختلفی: چن[4] و همکاران ۲۰۰۴، اکسی‌اوگلو[5] و همکاران ۲۰۰۵، وانگ[6] و همکاران ۲۰۰۶، خان[7] و همکاران ۲۰۰۷ از روش گزینش پیش‌رونده به عنوان یک ابزار برای پیش‌بینی و مدل‌سازی پدیده‌ها استفاده نمودند (۲۱). در روش رگرسیون گام‌به‌گام پیش‌رونده که براساس مدل رگرسیون خطی بنا نهاده شده است، جهت قضاوت این‌که در هر مرحله آیا یک متغیر مستقل جدید باید به مدل افزوده شود یا خیر یک سطح α انتخاب می‌کنند، که در این مطالعه سطح معنی‌داری α برابر ۰/۰۵ انتخاب شده‌است. این روش با این فرض شروع می‌شود که هیچ متغیر مستقلی در مدل حضور ندارد و فقط عرض از مبدأ وجود دارد. اولین متغیر مستقل که برای ورود به معادله انتخاب می‌شود آن است که بزرگ‌ترین همبستگی ساده را با متغیر وابسته‌ی Y دارد. سپس یک متغیر مستقل برای ورود به مدل برگزیده می‌شود. به بیان دیگر مدل رگرسیون ساده را برای هریک از پارامترهای مستقل، برازش داده و برای هر مدل رگرسیون ساده آماری مقدار آماره‌ی F را محاسبه می‌کنند. آن متغیر مستقلی که دارای مقدار F بزرگ‌تری است، برای افزودن به مدل انتخاب می‌شود. چنان‌چه F مربوط به این متغیر مستقل انتخاب شده از F_α (1,n-2) بزرگ‌تر باشد، آن‌گاه متغیر مستقل مربوطه به مدل افزوده می‌شود، در غیر این صورت افزودن این متغیر مستقل به مدل چندان مفید نخواهد بود. سرانجام مجموعه‌ای که با افزودن پارامتر مستقل دیگر در سطح α افزایش

1- Stepwise
2- Forward selection
3- Backward elimination
4- Chen
5- Eksioglu
6- Wang
7- Khan

معنی‌داری در آماره‌ی F آن ایجاد نشود، به عنوان بهترین ترکیب ورودی برای مدل‌سازی انتخاب می‌گردد (۱).

آزمون گاماتست

گاماتست یک ابزار مدل‌سازی غیرخطی است که به کمک آن می‌توان ترکیب مناسب از پارامترهای ورودی برای مدل‌سازی داده‌های خروجی و ایجاد یک مدل هموار را بررسی نمود. هم‌چنین یک ابزار توسعه یافته برای تخمین میانگین مربعات خطای حاصل از مدل‌سازی پدیده‌های مختلف با استفاده از مجموعه داده‌های مشاهداتی از آن پدیده است. با فرض آن‌که مجموعه‌ای از داده‌های ورودی x_i و خروجی y_i مشاهده شده از پدیده‌ای به صورت رابطه‌ی ۲ در اختیار باشد.

$$\{(x_i, y_i), 1 \le i \le M\} \qquad (۲)$$

در صورتی که با توجه به مجموعه ورودی x_i، خروجی y_i بدست آید و M نشان‌دهنده‌ی تعداد سری مجموعه‌ی مشاهده شده از پدیده‌ی مورد بررسی باشد، آن‌گاه گاماتست با فرض عدم قطعیت و پذیرش مقدار خطا در بدست آوردن خروجی از روی ورودی بخاطر پیچیدگی و غیرخطی بودن پدیده‌های مدل‌سازی این خطا را به صورت رابطه‌ی ۳ بین مجموعه داده‌های ورودی و خروجی با تابع f نشان می‌دهد.

$$y = f(x_1, x_2, \ldots, x_m) + r \qquad (۳)$$

که در آن f نمایانگر تابع همواری است که برای مدل‌سازی داده‌ها استفاده می‌شود و r نشان دهنده‌ی متغیر تصادفی است که برای نمایش خطا بکار می‌رود. با فرض آن‌که میانگین توزیعی که r از آن پیروی می‌کند برابر صفر است (هر مقدار ثابت دیگر را نیز می‌توان برای میانگین تابع f در نظر گرفت) و فرض دیگر آن‌که واریانس خطا کران‌دار است، مدل فوق به مدلی با مشتقات جزئی درجه اول محدود می‌شود و گاماتست نشان دهنده‌ی آن بخش از واریانس داده‌های خروجی است که نمی‌تواند توسط مدل هموار محاسبه گردد.

گاماتست بر اساس N[i,k] می‌باشد که در آن p بیانگر نزدیک‌ترین همسایگی است. مقادیر این آزمون برای مجموعه داده‌های ورودی و خروجی بر اساس روابط ۴ و ۵ بدست می‌آید که $|\cdot\cdot\cdot|$ فاصله‌ی اقلیدسی است.

$$\delta_M(k) = \frac{1}{M}\sum_{i=1}^{M}\left|x_{N(i,k)} - x_i\right|^2 \qquad 1 \le k \le p \qquad (۴)$$

$$\gamma_M(k) = \frac{1}{2M}\sum_{i=1}^{M}\left|y_{N(i,k)} - y_i\right|^2 \qquad 1 \le k \le p \qquad (۵)$$

با ایجاد رابطه‌ی رگرسیون خطی بین P مجموعه‌ی $(\gamma_M(k), \delta_M(k))$، مقدار آماره گاما برابر عرض از مبدا خط رگرسیون ایجاد شده‌است که معادله‌ی این خط در رابطه‌ی ۶ آورده شده‌است.

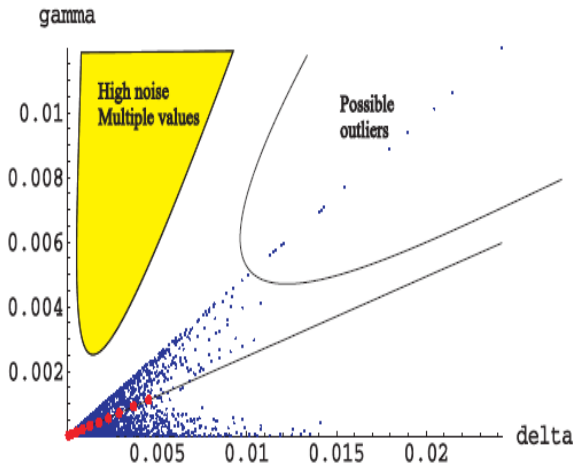
$$\gamma = A\delta + \Gamma \qquad (۶)$$

از روی خط رگرسیون بیان شده در رابطه‌ی ۶ و نمودار پراکنش گاماتست در شکل ۱ می‌توان اطلاعات مفیدی بدست آورد: نخست آن که عرض از مبدأ خط مذکور نشان دهنده‌ی مقدار گاماتست است که بیانگر آن بخش از واریانس داده‌های خروجی است که نمی‌تواند بوسیله‌ی مدل برآورد گردد. دوم آن که شیب خط رگرسیون نشان دهنده‌ی پیچیدگی مدلی است که از روی مجموعه‌ی داده‌های ورودی و خروجی ساخته می‌شود و این شیب هرچه تندتر باشد نشان دهنده‌ی پیچیدگی بیش‌تر مدل است. یکی دیگر از معیارهای مهم که با استفاده از این آزمون می‌توان بدست آورد معیار بدون بعد V_{ratio} است که دارای مقادیری بین بازه۰ و ۱ است و هرچه این مقدار به صفر نزدیک‌تر باشد نمایانگر دقت بالای مدل برای یافتن خروجی‌های مطلوب از ورودی‌هاست. در واقع اگر مقدار V_{ratio} از عدد یک کم شود مقدار ضریب تبیین را نشان می‌دهد (۲۵).

اگر فرض شود N پارامتر ورودی بر وقوع پدیده‌ای مؤثر باشند تعداد 2^N -1 ترکیب معنی‌دار از پارامترهای ورودی بوجود می‌آید که برای مدل‌سازی این پدیده با استفاده از ANNs بررسی تک‌تک ترکیبات ایجاد شده برای یافتن بهترین ترکیب کاری بسیار وقت‌گیر و خسته‌کننده است. بنابراین هنگامی که عوامل مؤثر بر پدیده‌ای بطور قابل ملاحظه‌ای زیاد باشد با استفاده از گاماتست می‌توان ترتیب میزان اهمیت پارامترهای ورودی و بهترین ترکیب از میان تمام ترکیب‌های ممکن را بدست آورد. هم‌چنین با استفاده از آزمون M تست می‌توان تعداد داده‌های ورودی که مقدار گاما را به حالت پایدار برای ایجاد مدل هموار می‌رساند، تعیین نمود (۱۹).

معیارهای آماری سنجش مقدار خطا

در این تحقیق به‌علت عدم وجود داده‌های لایسیمتری در ایستگاه مورد مطالعه، به منظور سنجش میزان کارایی و عملکرد مدل‌های مختلف، مقادیر تبخیر و تعرق روزانه‌ی محاسبه شده از مدل‌های مختلف (E_{Modeli}) با مقادیر برآوردی توسط روش فائو پنمن–مانتیث (E_{PM56}) به عنوان روش مرجع، از طریق معیارهای آماری ضریب تبیین (R^2)، ریشه میانگین مربعات خطا (RMSE) و MBE مقایسه می‌گردند.



شکل ۱- نمودار پراکنش گاماتست

$$MBE = \frac{\sum_{i=1}^{n}\left(E_{Modeli} - E_{PM\,56i}\right)}{n} \qquad (۷)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}\left(E_{Modeli} - E_{PM\,56i}\right)^2}{n}} \qquad (۸)$$

جاکوویدز معتقد است که معیارهای آماری RMSE و MBE به تنهایی برای انتخاب بهترین مدل ET_o کافی نمی‌باشند. بنابراین برای ارزیابی دقت مدل‌ها و مقایسه نسبی نتایج با مقادیر محاسبه شده از روش E_{PM56} از آزمون زیر که توسط جاکوویدز پیشنهاد گردیده است، نیز استفاده شد. بنابراین، علاوه بر دو معیار RMSE و MBE که عموماً برای مقایسه مدل‌های تبخیر و تعرق استفاده می‌شوند، از معیار دیگری به نام t که ترکیبی از دو معیار ذکرشده است، برای مقایسه‌ی بیشتر نیز استفاده شده‌است (۷).

$$t = \sqrt{\frac{(n-1)(MBE)^2}{\left(RMSE^2 - MBE^2\right)}} \qquad (۹)$$

در رابطه‌ی ۹، t معیار جاکوویدز و n تعداد مشاهدات می‌باشد و مقادیر کمتر معیار جاکویدز نشانگر کارایی بالاتر مدل در مدل‌سازی است.

مدل‌های مختلف پیش‌پردازش پارامترهای ورودی

در این تحقیق با پیش‌پردازش پارامترهای ورودی توسط روش‌های رگرسیون گام‌به‌گام و گاماتست، سه مدل مختلف جهت انتخاب بهترین ترکیب از میان پارامترهای ورودی، تعیین شد. اولین مدل که با استفاده از رگرسیون گام‌به‌گام بدست آمد ANN -FS نام دارد. مدل دوم که بر پایه‌ی پارامترهای مؤثر بدست آمده از گاماتست ساخته شد ANN-GT نام دارد. مدل سوم که بدون انجام هیچ‌گونه پیش‌پردازشی بر روی پارامترهای ورودی، همه‌ی آن‌ها را شامل می‌شود مدل ANN نام دارد. با تعیین پارامترهای مؤثر و ترکیب بهینه در هر روش، به کمک شبکه‌ی عصبی MLP به مدل‌سازی ET_o اقدام و سپس با معیارهای مختلف آماری سنجش مقدار خطا، به ارزیابی مدل‌های ایجاد شده جهت تخمین تبخیر و تعرق روزانه پرداخته شد.

نتایج و بحث

نتایج آزمون رگرسیون گام‌به‌گام پیش‌رونده

بر اساس نتایج این آزمون در مرحله‌ی نخست تابش‌برون‌زمینی مهم‌ترین پارامتری است که برای ورود به مدل انتخاب می‌شود. زیرا مطابق با نتایج جدول ۲ که در آن مقادیر ضریب همبستگی ساده (دو به دو) پارامترهای هواشناسی آمده‌است، پارامتر تابش‌برون زمینی دارای بیشترین مقدار ضریب همبستگی با تبخیر و تعرق به مقدار ۸۸/۳۹ درصد می‌باشد. بنابراین وقتی به مدل اضافه می‌شود مقدار ضریب تبیین را به ۷۸/۱۲ درصد می‌رساند. در گام بعدی پارامتر دمای حداکثر با بیشترین مقدار F نسبت به سایر پارامترهای مستقل باقی‌مانده، به مدل اضافه می‌شود که باعث افزایش این ضریب به عدد ۸۷/۸۳ درصد می‌شود. سپس از میان مجموعه پارامترهای باقی‌مانده پارامتری که می‌تواند با بیشترین مقدار F و معنی‌داری در سطح α

برابر ۰/۰۵ به مدل اضافه شود، سرعت باد است که ضریب تبیین مدل را به ۹۴/۳ درصد ارتقاء می‌دهد و به همین ترتیب پارامترهایی که در گام‌های بعدی می‌توانند در سطح α باعث ایجاد معناداری در مقدار آماره‌ی F شوند، پارامترهای ساعات آفتابی، دمای متوسط و رطوبت نسبی می‌باشند و در نهایت اضافه نمودن پارامتر دمای حداقل نمی‌تواند در سطح ۰/۰۵ سبب ایجاد معنی‌داری در مقدار آماره‌ی F شود. بنابراین براساس نتایج جدول ۳، مدل حاصل از روش رگرسیون گام به‌گام به روش پیشرو از تمام پارامترهای ورودی به جز پارامتر دمای حداقل برای مدل‌سازی تبخیر و تعرق استفاده می‌نماید. نتایج گام به گام مدل‌های ایجاد شده در هر مرحله به همراه مقادیر p-value، آماره‌ی F و مقدار ضریب تبیین برای این آزمون در جدول ۳ ارائه شده‌است.

نتایج گاماتست

با استفاده از روش گاماتست برای پیش‌پردازش پارامترها می‌توان ترتیب میزان اهمیت پارامترهای ورودی، بهترین ترکیب از میان تمام ترکیب‌های ممکن و تعداد داده‌های مورد نیاز برای ایجاد یک مدل هموار جهت ورود به شبکه‌ی عصبی مصنوعی را بدست آورد.

اهمیت پارامترهای ورودی

در این مطالعه ترکیبات مختلف ورودی مورد آزمون قرار گرفته‌اند تا تاثیر هر پارامتر بر مقدار تبخیر و تعرق سنجیده شود. برای مشخص نمودن ترتیب اهمیت پارامترهای ورودی ابتدا آزمون GT را برای ترکیبی که همه‌ی پارامترهای ورودی در آن دخالت دارند انجام داده و سپس در گام بعدی یکی از پارامترها را به دلخواه از مجموعه‌ی اولیه حذف و این بار گاماتست را در این مطالعه با ۶ پارامتر باقی‌مانده محاسبه می‌نمایند و سپس در گام بعدی متغیر حذف شده را به مجموعه پارامترهای اصلی اضافه نموده و پارامتر دیگری را حذف می‌نمایند. این فرآیند برای تمام متغیرها به ترتیب انجام می‌گردد و هر بار مقدار آماره‌ی گاماتست محاسبه می‌شود. در این فرآیند با حذف پارامتر تأثیرگذار مقدار آماره‌ی GT افزایش می‌یابد و با حذف پارامتری که از اهمیت کمتری برخوردار است مقدار این آماره کاهش می‌یابد. نتایج این آزمون برای تبخیر و تعرق روزانه در جدول ۴ آورده شده است. نتایج جدول ۴ بیانگر آن است که با حذف پارامتر سرعت باد، آماره‌ی گاما دارای بیشترین مقدار می‌گردد. بنابراین با توجه به این آزمون سرعت باد مؤثرترین پارامتر بر ET_o در ایستگاه سینوپتیک شیراز می‌باشد و پارامترهای تابش برون‌زمینی، ساعات آفتابی، رطوبت نسبی، دمای متوسط، دمای حداقل و دمای حداکثر به ترتیب در مرتبه‌های بعدی اهمیت قرار دارند. هم‌چنین با رسم نمودار منحنی‌های مجانب برای ترکیبی که همه‌ی پارامترها در آن دخالت دارند و ۷ ترکیب دیگر ذکر شده در جدول ۴، که در آن‌ها در هر مرحله یکی از پارامترهای ورودی به دلخواه حذف گردیده‌است، می‌توان به ترتیب اهمیت پارامترهای ورودی دست یافت. نتایج حاصل از منحنی‌های مجانب مقادیر گاما برای ۸ ترکیب یاد شده در شکل ۲ ارائه شده‌است. مطابق منحنی‌های مجانب رسم شده در این شکل با حذف پارامتر سرعت باد منحنی مجانب این ترکیب در فاصله‌ی بالاتری نسبت به سایر مجانب‌ها قرار می‌گیرد. بنابراین از تأثیرگذاری بیشتری نسبت به سایر پارامترها برخوردار است.

تعیین بهترین ترکیب

انتخاب ترکیب مناسب از پارامترهای ورودی یکی از مهم‌ترین مراحل ساخت و طراحی هرگونه مدل‌سازی ریاضی و هوشمند است. اگر فرض شود N پارامتر ورودی بر وقوع پدیده‌ای مؤثر باشند 2^N- 1 ترکیب معنی‌دار از پارامترهای ورودی بوجود می‌آید.

جدول ۲- مقادیر ضریب همبستگی ساده (دو به دو) پارامترهای هواشناسی

پارامتر	دمای حداکثر	دمای حداقل	دمای متوسط	رطوبت نسبی	ساعات آفتابی	سرعت باد	تابش‌برون زمینی	تبخیرو تعرق
تبخیر وتعرق	۰/۸۶۹	۰/۸۵۱	۰/۸۷۶	-۰/۶۴۱	۰/۶۱۴	۰/۵۲۴	۰/۸۸۴	۱
تابش‌برون‌زمینی	۰/۷۵۰	۰/۷۵۸	۰/۷۵۲	-۰/۵۹۱	۰/۴۵۱	۰/۳۶۶	۱	
سرعت باد	۰/۲۰۱	۰/۲۸۷	۰/۲۴۴	-۰/۲۲۲	۰/۱۰۸	۱		
ساعات آفتابی	۰/۵۹۹	۰/۵۱۷	۰/۳۹۶	-۰/۷۶۱	۱			
رطوبت نسبی	-۰/۸۱۱	-۰/۷۵۴	-۰/۶۵۵	۱				
دمای متوسط	۰/۹۸۶	۰/۹۷۸	۱					
دمای حداقل	۰/۹۲۹	۱						
دمای حداکثر	۱							

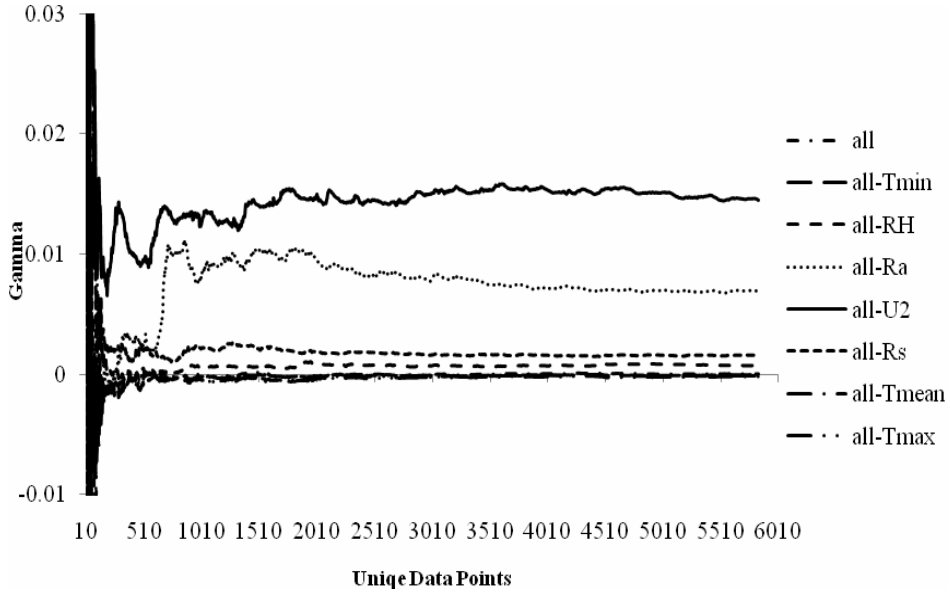
جدول ۳- نتایج آزمون رگرسیون گام‌به‌گام پیش‌رونده

پارامترهای ورودی مدل	P-value	آماره F	ضریب تبیین
تابش برون‌زمینی	۰/۰۰۰۰	۲۰۸۱۹/۳۲***	۷۸/۱۲
تابش برون‌زمینی، دمای حداکثر	۰/۰۰۰۰	۴۶۴۹/۳۹***	۸۷/۸۳
تابش برون‌زمینی، دمای حداکثر، سرعت باد	۰/۰۰۰۰	۶۶۲۰/۳۱***	۹۴/۳۰
تابش برون‌زمینی، دمای حداکثر، سرعت باد، ساعات آفتابی	۰/۰۰۰۰	۲۰۱۰/۰۶***	۹۵/۷۶
تابش برون‌زمینی، دمای حداکثر، سرعت باد، ساعات آفتابی، دمای متوسط	۰/۰۰۰۰	۵۱۴/۶۵***	۹۶/۱۰
تابش برون‌زمینی، دمای حداکثر، سرعت باد، ساعات آفتابی، دمای متوسط، رطوبت نسبی	۰/۰۰۰۰	۶۰/۷۶***	۹۶/۱۴

*** معنی داری در سطح α (۰/۰۵)

جدول ٤- نتایج مقادیر آماره‌ی گاما برای ترکیب پارامترهای مختلف ورودی

مقادیر گاما	ترکیب پارامترهای ورودی	مقادیر گاما	ترکیب پارامترهای ورودی
۰/۰۱۴۵۴	T_{min}, T_{max}, T_{mean}, R_a, R_s, RH	-۰/۰۰۰۱۵	T_{min}, T_{max}, T_{mean}, R_a, R_s, RH, U_2
-۰/۰۰۰۱۳	T_{min}, T_{max}, R_a, R_s, RH, U_2	۰/۰۰۱۵۹	T_{min}, T_{max}, T_{mean}, R_a, RH, U_2
۰/۰۰۰۰۱۱	T_{min}, T_{mean}, R_a, R_s, RH, U_2	۰/۰۰۶۸۹	T_{min}, T_{max}, T_{mean}, R_s, RH, U_2
-۰/۰۰۰۰۲	T_{max}, T_{mean}, R_a, R_s, RH, U_2	۰/۰۰۰۷۲	T_{min}, T_{max}, T_{mean}, R_a, R_s, U_2



شکل ۲ - منحنی‌های مجانب مقادیر گاما در حالت حذف هریک از پارامترها ورودی

جدول ۵- نتایج مقادیر آماره‌های گاما، گرادیانت، خطای استاندارد و v_{ratio} برای هفت ترکیب منتخب

پارامتر	All-T_{max}	All-T_{min}	All-$T_{max,,min}$	All-$T_{min,,mean}$	All-$T_{max,,mean}$	All-T_{mean}	All
مقدار گاما	۰/۰۰۰۰۱	-۰/۰۰۰۰۲	۰/۰۰۰۰۶	۰/۰۰۰۳۹	۰/۰۰۰۵۳	-۰/۰۰۰۱۴	-۰/۰۰۰۱۵
گرادیانت	۰/۰۳۵۵	۰/۰۳۴۴	۰/۰۴۱۴	۰/۰۴۲۲	۰/۰۴۰۹	۰/۰۳۳۷	۰/۰۳۰۹
خطای استاندارد	۰/۰۰۰۰۵۶	۰/۰۰۰۰۴۴	۰/۰۰۰۰۵۸	۰/۰۰۰۰۷۴	۰/۰۰۰۰۷۷	۰/۰۰۰۰۴۸	۰/۰۰۰۰۳۹
V_{ratio}	۰/۰۰۰۰۴	۰/۰۰۰۰۹	۰/۰۰۰۲۲	۰/۰۰۱۵۷	۰/۰۰۲۱۱	۰/۰۰۰۵۴	۰/۰۰۰۶۳

برای مثال در این تحقیق با در نظرگـرفتن هفـت پـارامتر ورودی مؤثر بر ET_o تعداد ۱۲۷ ترکیب مختلف ایجاد می‌شود که برای مدل-

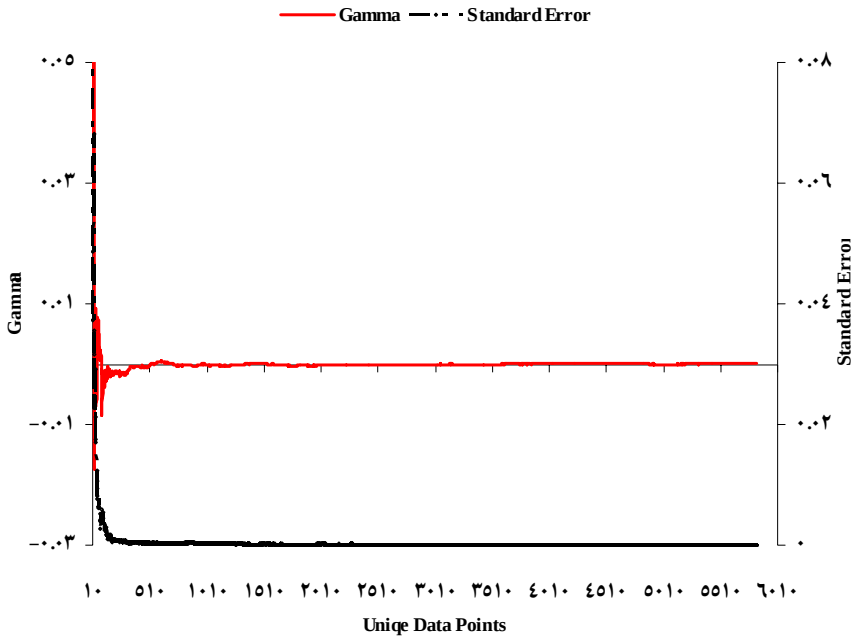
سازی این پدیده با استفاده از ANNs بررسی تک‌تک ترکیبات ایجاد شده برای یافتن بهترین ترکیب کاری بسیار وقت‌گیر و خسته‌کننده است. بنابراین با استفاده از این آزمون با بررسی ۱۲۷ ترکیب ممکن ایجاد شده نتایج حاصل از آن برای هفت ترکیب منتخب که دارای کمترین مقدار گاما هستند در جدول۵ آورده شده است. براساس نتایج جدول ۵ و شکل ۲ ترکیبی که شامل همه‌ی پارامترهای ورودی به جز دمای حداکثر باشد با کم‌ترین مقدار آماره‌ی گاما نسبت به سایر ترکیب‌ها و پائین‌ترین منحنی مجانب که دارای منحنی مجانب تقریباً همواری است، به عنوان بهترین ترکیب ورودی انتخاب می‌شود.

تعیین تعداد داده‌های مورد نیاز با استفاده از آزمون M تست

یکی از چالش‌های اصلی در مدل‌سازی پدیده‌ها، ارزیابی اطلاعات موجود و کفایت داده‌ها است. به بیان دیگر چه تعداد الگو از مجموعه الگوهای ورودی باید برای مدل‌سازی استفاده شود؟ اگر چه به طور کلی با افزودن اطلاعات بیش‌تر عملکرد مدل در مرحله‌ی کالیبراسیون بهبود می‌یابد ولی باید به این نکته توجه داشت که افزودن اطلاعات بیش‌تر به چه میزان می‌تواند عملکرد مدل را بهبود ببخشد؟ (۲۳). علاوه برآن با استفاده از نمودار M تست می‌توان تعداد داده‌های معنی‌دار برای ایجاد یک مدل هموار برای پیش‌بینی تبخیر و تعرق روزانه و چگونگی تغییرات خطای استاندارد و مقدار آماره‌ی گاما با افزایش تعداد داده‌ها را نشان داد. با توجه به شکل ۳ می‌توان مشاهده نمود که منحنی‌های مجانب آماره‌های گاما و خطای استاندارد تقریبا در اطراف نقطه‌ی ۸۰۰ به حالت پایدار می‌رسد که با افزایش تعداد داده‌ها مجانب‌های گاما و خطای استاندارد تغییرات چندانی به خود نمی‌بینند. بنابراین تعداد داده‌های معنی‌دار برای آموزش شبکه‌های عصبی مصنوعی ۸۰۰ مجموعه انتخاب می‌شود.

نتایج پیش‌پردازش پارامترهای ورودی در مدل‌های مختلف

بر اساس نتایج پارامترهای مؤثر بدست آمده از روش‌های رگرسیون گام‌به‌گام و گاماتست بهترین ترکیب از میان پارامترهای ورودی برای سه مدل مختلف برای ورود به شبکه‌ی عصبی MLP، مشخص گردید که نتایج آن برای مدل‌های مختلف در جدول ۶ آمده است. مطابق با نتایج این جدول مدل ANN-FS برای مدل‌سازی ET_o از ترکیب تمام پارامترهای ورودی به جزء دمای حداقل استفاده می‌نماید. بر پایه‌ی بهترین ترکیب بدست آمده از گاماتست، مدل ANN-GT نیز از ترکیب تمام پارامترهای ورودی به جزء دمای حداکثر استفاده و مدل ANN سوم نیز بدون انجام هیچ‌گونه پیش‌پردازشی بر روی پارامترهای ورودی، از ترکیب همه‌ی آن‌ها استفاده می‌کند.



شکل ۳ – نمودار آزمون M تست مقادیر گاما و خطای استاندارد برای مدل برگزیده در گاماتست

جدول ٦- مدل‌های برگزیده و پارامترهای ورودی در هر مدل

نام مدل	ترکیب پارامترهای ورودی
ANN-FS	T_{max}, T_{mean}, R_a, R_s, RH, U_2
ANN-GT	T_{min}, T_{mean}, R_a, R_s, RH, U_2
ANN	T_{min}, T_{max}, T_{mean}, R_a, R_s, RH, U_2

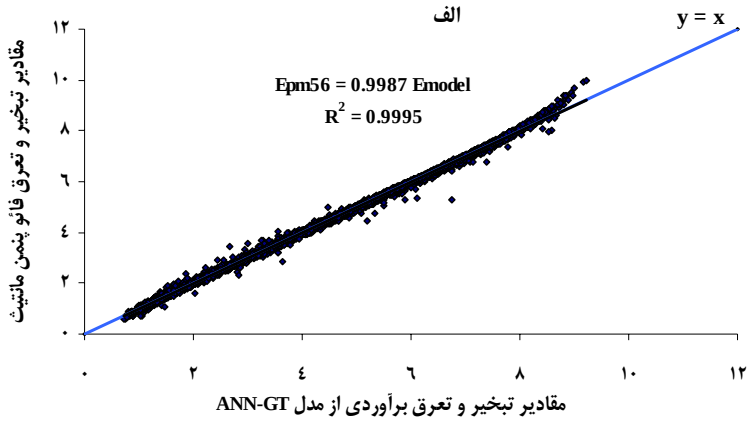
نتایج شبکه‌های عصبی مصنوعی برای داده‌های آزمون

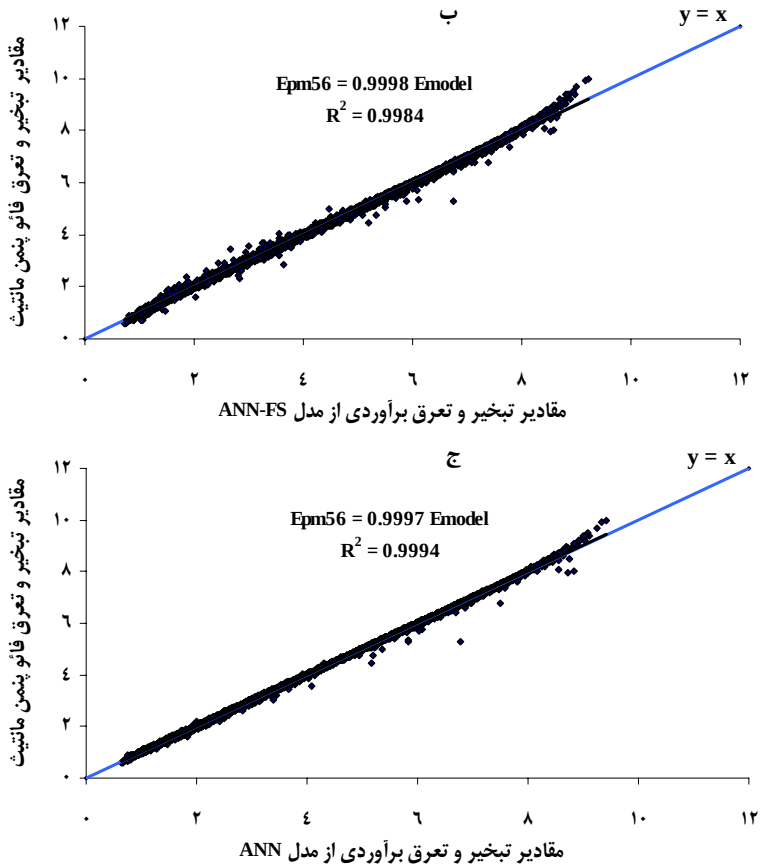
مطابق نتایج حاصل از پیش‌پردازش پارامترهای ورودی در جدول ۶ برای آموزش هریک از سه مدل تعریف شده با استفاده از شبکه‌ی عصبی MLP، در لایه‌ی ورودی از ترکیب پارامترهای هواشناسی مختلف استفاده شد. بدین منظور از ۵۸۳۲ رکورد ثبت شده این ایستگاه در سال‌های ۱۹۹۲ الی ۲۰۰۷ با توجه به آزمون M تست تعداد ۸۰۰ رکورد برای آموزش و تعداد ۵۰۳۳ رکورد باقیمانده برای آزمون شبکه‌های عصبی استفاده شده‌است. برای این کار در لایه‌ی پنهان از یک لایه‌ی مخفی با تعداد نرون‌های متفاوتی استفاده شده و تعداد بهینه‌ی آن‌ها برای حداقل نمودن خطا تعیین گردیده است. روند کار با تعداد نرون‌های کم آغاز و افزودن نرون‌های اضافی تا زمانی ادامه داشت که افزایش نرون‌های بیش‌تر تأثیری در بهبود خطا نداشته باشند. بدین منظور از ۲ تا ۲۰ نرون استفاده گردید. هم‌چنین در لایه‌ی پنهان از توابع محرک تانژانت سیگموئید، لوگ سیگموئید با روش پس‌انتشار خطا و الگوریتم آموزشی لونبرگ–مارکوارت مورد استفاده قرار گرفت. تعداد نرون‌های لایه‌ی پنهان، توابع محرک به همراه نتایج حاصل از مقایسه‌ی سه مدل ANN-GT، ANN-FS و ANN با معیارهای آماری مختلف سنجش خطا در جدول ۷ آورده شده است. براساس نتایج این جدول هر سه مدل از دقت بالایی جهت تخمین ET_o روزانه برخوردارند و از میان آن‌ها، مدل ANN-GT با مقدار ضریب تبیین (R^2) ۰/۹۹۹۵، آماره جاکویدز (t) ۶/۱۹۲ و ریشه میانگین مربعات خطا (RMSE) ۰/۰۴۸۳ نسبت به دو مدل دیگر ANN-FS و ANN به ترتیب با مقادیر R^2 برابر با ۰/۹۹۸۴ و ۰/۹۹۹۴، t برابر با ۶/۷۴۹ و ۶/۸۴۸ و RMSE برابر با ۰/۰۸۷۴ و ۰/۰۵۴۸ از دقت بالاتری برخوردار است.

هم‌چنین نتایج حاصل از سه مدل برای مقادیر فائو پنمن مانتیث به عنوان داده‌های واقعی و مقادیر بدست آمده از شبیه‌سازی توسط شبکه‌ی MLP، روی خط برازش یافته و خط ۴۵ درجه (۱:۱) در شکل ۴ نشان‌داده شده‌است. براساس نتایج این شکل ضریب شیب خط بدون عرض از مبدا برای هریک از سه مدل ANN-GT، ANN-FS و ANN به ترتیب برابر ۰/۹۹۸۷، ۰/۹۹۸۸ و ۰/۹۹۹۷ می‌باشد که تقریبا هر سه مدل دارای شیبی مساوی و برابر با شیب خط ۴۵ درجه می‌باشند.

جدول ۷- مقایسه‌ی سه مدل ANN-GT، ANN-FS و ANN با معیارهای آماری مختلف سنجش خطا

مدل	ساختار	تابع محرک	R^2	RMSE	MBE	t
ANN-FS	۶–۵–۱	لوگ سیگموئید	۰/۹۹۸۴	۰/۰۸۷۳۹	۰/۰۰۸۴	۶/۸۴۹
ANN-GT	۶–۸–۱	تانژانت سیگموئید	۰/۹۹۹۵	۰/۰۴۸۳۲	۰/۰۰۴۲	۶/۱۹۲
ANN	۷–۸–۱	تانژانت سیگموئید	۰/۹۹۹۴	۰/۰۵۴۷۵	۰/۰۰۵۲	۶/۷۶۸





شکل ٤- نتایج مدل‌های (الف) ANN-GT، (ب) ANN-FS و (ج) ANN روی خط ٤٥ درجه (۱:۱) جهت تخمین ET_0 (میلیمتر بر روز) در ایستگاه سینوپتیک شیراز

نتیجه‌گیری کلی

انتخاب ترکیب مناسب از پارامترهای ورودی یکی از مهم‌ترین مراحل ساخت و طراحی هرگونه مدل‌سازی ریاضی و هوشمند است. در این تحقیق از روش رگرسیون گام‌به‌گام و یک ابزار جدید به نام گاماتست برای پیش‌پردازش پارامترهای ورودی و انتخاب ترکیب بهینه از پارامترهای ورودی جهت شبیه‌سازی تبخیر و تعرق روزانه به کمک شبکه‌ی عصبی MLP استفاده شد. آن‌گاه با مرجع قراردادن روش فائوپنمن مانتیث، عملکرد مدل‌های مختلف جهت تخمین تبخیر روزانه مورد ارزیابی و بررسی قرار گرفت. نتایج ارزیابی سه مدل نشان داد که هر سه مدل از دقت بالایی برای شبیه‌سازی تبخیر و تعرق روزانه برخوردارند. ضمن آن‌که برتری مدل ANN-GT نسبت به مدل ANN و نیز برتری مدل‌های ANN-GT و مدل ANN نسبت به مدل ANN-FS چندان محسوس نمی‌باشد.

نتایج پیش‌پردازش پارامترهای ورودی روش‌های FS و GT جهت مدل‌سازی نشان داد که افزودن پارامترهای دمای حداقل و دمای حداکثر در بهبود عملکرد مدل شبیه‌سازی تأثیر چندانی ندارند و این واقعیت می‌تواند ناشی از همبستگی بالا بین پارامترهای دمای حداقل و دمای حداکثر با دمای میانگین باشد که باعث ایجاد هم‌خطی و نقض استقلال پارامترهای ورودی مدل می‌شود. بنابراین مدل بهینه در این دو روش به خوبی توانسته است با حذف یکی از این پارامترها مانع ایجاد هم‌خطی شود.

همان‌طور که در بخش پایانی مقدمه ذکر شد، اهداف این تحقیق پاسخ به چند سوال اساسی در ارتباط با مدل‌سازی تبخیر و تعرق با استفاده از شبکه‌ی عصبی از جمله: کدام یک از داده‌های ورودی به مدل از اهمیت بیشتری برخوردار است؟ چه ترکیبی از پارامترهای ورودی باید به مدل وارد شود؟ چه تعداد داده برای مدل‌سازی تبخیر و تعرق با توجه به مجموعه‌ی ورودی مورد نیاز است؟ تعریف شده بود که در این تحقیق روش‌های FS و GT برای پاسخ به سوالات فوق و ارائه‌ی یک روش قاعده‌مند برای کاهش مراحل سعی و خطا بکارگرفته شد. نتایج بکارگیری روش‌های پیش‌پردازش فوق نشان داد که روش FS با مشخص نمودن اهمیت نسبی پیوستن پارامترها به مدل در هرگام و تعیین مدل بهینه‌ی نهایی در سطح معنی‌داری ۰/۰۵ در پاسخ به دو پرسش از پرسش‌های مطرح شده می‌توانند به عنوان

روشی مؤثر بکار رود. ضمن آن‌که آزمون گاماتست به عنوان یک الگوی نو ضمن پاسخ به دو پرسش قبلی با توانایی تعیین تعداد الگوهای لازم برای آموزش شبکه جهت ایجاد یک مدل هموار با استفاده از آزمون M تست (در این تحقیق ۸۰۰ الگوی ورودی) می‌تواند به عنوان یک روش کارامد برای پیش‌پردازش پارامترهای ورودی جهت مدل‌سازی تبخیر و تعرق با استفاده از شبکه MLP بکارگرفته شود.

بنابراین به‌نظر می‌رسد ارائه و بکارگیری الگوهای نو و تکنیک‌های پیشرفته جهت پیش‌پردازش پارامترهای ورودی می‌تواند موجب کاهش مراحل سعی و خطا در یافتن ترکیب مناسب ورودی و ایجاد تحول در مدل‌سازی پدیده‌ها با استفاده از روش‌های هوشمند گردد.

منابع

۱- بازرگان لاری ع. ۱۳۸۴. رگرسیون خطی کاربردی. چاپ اول. انتشارات مرکز نشر دانشگاه شیراز، ۳۲۹ ص.

۲- جهانبخش‌اصل س.، موحددانش ع. و مولوی ا. ۱۳۸۰. تحلیل مدل‌های برآورد تبخیر و تعرق برای ایستگاه هواشناسی تبریز. مجله دانش کشاورزی ۱۱ : ۶۶-۵۱.

۳- رضایی ع. و سلطانی ا. ۱۳۸۲. مقدمه‌ای بر تحلیل رگرسیون کاربردی. دانشگاه صنعتی اصفهان، مرکز نشر،۲۹۴ ص

۴- زارع ابیانه ح.، بیات ورکشی م.، معروفی ص. و امیری چایجان ر. ۱۳۸۹. ارزیابی سیستم‌های هوشمند عصبی در کاهش پارامترهای تخمین تبخیر و تعرق گیاه مرجع. مجله آب و خاک (علوم و صنایع کشاورزی)، ۲۴ (۲) : ۳۰۵-۲۹۷.

۵- زارع ابیانه ح.، قاسمی ع.، بیات ورکشی م. و معروفی ص. ۱۳۸۸. ارزیابی دقت شبکه عصبی مصنوعی در پیش‌بینی تبخیر و تعرق گیاه سیر با استفاده از داده‌های لایسیمتری در منطقه همدان. مجله آب و خاک (علوم و صنایع کشاورزی)، ۲۳ (۳) : ۱۸۵-۱۷۶.

۶- علیزاده ا. و کمالی غ. ۱۳۸۶. نیاز آبی گیاهان در ایران. چاپ اول. موسسه‌ی چاپ و انتشارات آستان قدس رضوی. ۲۲۸ ص.

۷- سبزی‌پرور ع. ا.، تفضلی ف.، زارع ابیانه ح.، بانژاد ح.، موسوی بایگی م.، غفوری م.، محسنی موحد، ا. و مریانجی، ز. ۱۳۸۷. مقایسه چند مدل برآورد تبخیر و تعرق گیاه مرجع در یک اقلیم سرد نیمه‌خشک، به منظور استفاده بهینه از مدل‌های تابش. مجله آب و خاک (علوم و صنایع کشاورزی)، ۲۲ (۲) : ۳۴۰-۳۲۷ .

۸- شایان نژاد م.، ساداتی نژاد ج. و فهمی ه. ۱۳۸۶. تعیین تبخیر و تعرق بالقوه با استفاده رگرسیون فازی. مجله تحقیقات منابع آب ایران، ۳: ۱۹-۹.

۹- کوچک‌زاده م. و بهمنی ع. ۱۳۸۴. ارزیابی عملکرد شبکه‌های عصبی مصنوعی در کاهش پارامترهای مورد نیاز، جهت برآورد تبخیر و تعرق مرجع. مجله علوم کشاورزی، ۱۱ : ۹۶-۸۷.

۱۰- موسوی بایگی م.، عرفانیان م. و سرمد م. ۱۳۸۸. استفاده از حداقل داده‌های هواشناسی برای برآورد تبخیر و تعرق گیاه مرجع و ارائه‌ی ضریب اصلاحی (مطالعه موردی: استان خراسان رضوی). مجله آب و خاک (علوم و صنایع کشاورزی)، ۲۳ (۱) : ۹۹-۹۱.

11- Ahmadi A., Han D., Karamouz M., and Remesan R. 2009. Input data selection for solar radiation estimation. Hydrological Processes, 23: 2754–2764.

12- Allen, R.G., Pereira, L.S., Raes, D., and Smith, M., 1998. Crop evapotranspiration. Guidelines for computing crop water requirements. FAO Irrigation and Drainage, Paper no. 56. FAO, Rome.

13- Jones A.J., Evans D., Margetts S., and Durrant P. 2002 .The Gamma Test. Chapter IX in Heuristic and Optimization for Knowledge Discovery. Edited by Ruhul Sarker, Hussein Abbass and Charles Newton. Idea Group Publishing, Hershey, PA. 27 pp.

14- Jones A.J., Margetts S., and Durrant P. 2000. The win Gamma User Guide. Department of computer science, university of Wales, 77pp. accessible at: http://users.cs.cf.ac.uk/.../Gamma%20Software/winGamma/winGamma.htm/

15- Jones A.J., Tsui A., and Oliveira A.G. 2002. Neural models of arbitrary chaotic systems. construction and the role of time delayed feedback in control and synchronization, Complexity Int;09.

16- Kumar M., Bandyopadhyay A., Raghuwanshi N. S., and Singh R. 2008. Comparative study of conventional and artificial neural network-based ETo estimation models. Irrigation Sciences 26:531–545.

17- Kisi O., and Ozturk, O. 2007. Adaptive neuro-fuzzy computing technique for evapotranspiration estimation. J. Irrig. Drain. Eng., ASCE 133 (4), 368–379.

18- McCulloch W.S., and Pitts W.A. 1943. Logical calculus of the ideas immanent in nervous activity. Bull Math Biophys ,5:115−33.

19- Moghaddamnia A., Ghafari Gousheh M., Piri J., Amin S., and Han D. 2009. Evaporation estimation using artificial neural networks and adaptive neuro-fuzzy inference system techniques. Advances in Water Resources. 32 : 88–97.
20- Noori R., Karbassi A., and Sabahi M.S .2009. Evaluation of PCA and gamma test techniques on AAN opration for weekly solid waste prediction, Journal of Enviromental Management.91: 767-771
21- Noori R., Hoshyaripour GH., Ashrafi KH., and Najdar Araabi B. 2009. Uncertainty analysis of developed ANN and ANFIS models in prediction of carbon monoxide daily concentration, Atmospheric Environment, 44:476-482.
22- Odhiambo L.O., Yoder R. E., Yoder D. C., and Hines J.W. 2001. Optimization of fuzzy evaporation model through neural training with input-output examples. Trans, ASAE, 44(6):1625-1633.
23- Peterson A.T., and Cohoon K.P. 1999. Sensitivity of distributional prediction algorithms to geographic data completeness. Ecological Modelling 117: 159–164.
24- Rahimi Khoob A. 2008. Artificial neural network estimation of reference evapotranspiration from pan evaporation in a semi-arid environment. Irrigation Sciences 27:35–39
25- Remesan R., Shamim M.A., and Han D. 2008. Model data selection using gamma test for daily solar radiation estimation. *Hydrological Processes* 22: 4301–4309.
26- Sudheer K.P., Gosain A.K., and Ramasastri K.S. 2003. Estimating actual evapotranspiration from limited climatic data using neural computing technique. J. Irrigation. Drainage. Engineering, 129: 214–218.
27- Venables W.N., Smith D.M., and the R Development Core Team. 2009. An Introduction to R. ISBN 3-900051-12-7, 100pp. accessible at: http://www.R-project.org/.

Journal of Water and Soil
Vol. 24, No. 3, July-Aug 2010, p. 610-624

نشریه آب و خاک
جلد ۲۴، شماره ۳، مرداد - شهریور ۱۳۸۹، ص.۶۲۴-۶۱۰

Evaluation Effect of Input Parameters Preprocessing in Artificial Neural Networks (Anns) by Using Stepwise Regression and Gamma Test Techniques for Fast Estimation of Daily Evapotranspiration

M. Ghabaei Sough[1] - A. Mosaedi[2*] - M. Hesam[3] - A. Hezarjaribi[3]

Abstract

Process of evapotranspiration (ET_o) is a major component of the hydrologic cycle that its accurate estimation plays an important role to achieve sustainable development in water balance, irrigation system design and planning and management of water resources. Being a function of different metrological parameters and their interactions, evapotranspiration is a complex, nonlinear phenomenon. Preprocessing input parameters to select appropriate combinations is complex when modeling nonlinear systems. Using these methods reduce steps by trial and error by recognizing most important parameters for modeling by intelligent methods. In this study, two methods of stepwise regression (FS) and gamma test (GT) were used for preprocessing input parameters in multi layer perceptron neural network (MLP) to estimate daily estimation of ET_o at Shiraz synoptic station. To evaluate the effect of Input parameters preprocessing in artificial neural networks using different statistical error criteria, ANN-FS and ANN-GT both with pre-processed parameters were compared against each other and also with ANN model without pre-processed parameters. The results showed that all three models have a high degree of accuracy to estimate daily ET_o. ANN-GT model represented a determination coefficient (R^2) of 0.9995 and the root mean square error (RMSE) of 0.0483. For ANN-FS and ANN models R^2 values were 0.9984 and 0.9994 respectively and $RMSE_s$ were 0.0874 and 0.0548 respectively. Although the accuracy of ANN-GT model was slightly greater than ANN, but the ability of determination of importance of input parameters, education and recognition of the best combination of input parameters with 800 data in this study, makes this model a useful tool for fast preprocessing input parameters to model evapotranspiration.

Keywords: Potential evapotranspiration, Artificial neural networks, Gamma test, Stepwise regression, Shiraz synoptic station

1,2,3,4- MSc. Student, Associate prof., and Assistant professors, Dept. of Water Eng. Gorgon University of Agricultural Sciences and Natural Resources, Respectively
(* - Corresponding author Email: mosaedi@yahoo.com)