

پیش‌پردازش پارامترهای ورودی به شبکه‌ی عصبی مصنوعی و سیستم استنتاج تطبیقی عصبی - فازی با استفاده از رگرسیون گام به گام و گاماتست به‌منظور تخمین تبخیر

محمد زمانیان^۱، روح‌اله فتاحی^۲، احمدرضا فتاحی^۳ و فرشته حسین‌پور^۴

تاریخ دریافت: ۹۰/۱۰/۱۹

تاریخ پذیرش: ۹۱/۰۹/۱۷

چکیده

فرایند تبخیر به‌علت نیاز به فاکتورهای اقلیمی مختلف و اثر متقابل این فاکتورها بر یکدیگر، یک پدیده غیرخطی و پیچیده است. یکی از مراحل پیچیده در مدل‌سازی غیرخطی، پیش‌پردازش پارامترهای ورودی برای انتخاب ترکیبی مناسب از آنها است. پیش‌پردازش داده‌ها سبب کاهش مراحل سعی و خطا و شناخت مهم‌ترین پارامترهای مؤثر بر پدیده‌ی مورد نظر به‌منظور مدل‌سازی با استفاده از روش‌های هوشمند می‌شود. در این پژوهش از دو روش رگرسیون گام به گام (FS) و گاماتست (GT) برای پیش‌پردازش پارامترهای ورودی به شبکه‌ی عصبی پرسپترون چندلایه و سیستم استنتاج تطبیقی عصبی - فازی برای تخمین تبخیر روزانه‌ی ایستگاه هواشناسی شهرکرد استفاده شده است. برای ارزیابی تأثیر پیش‌پردازش پارامترهای ورودی با استفاده از معیارهای مختلف آماری سنجش خطا به مقایسه‌ی چهار مدل ANN-FS، ANN-GT، ANFIS-FS و ANFIS-GT (با پارامترهای پیش‌پردازش شده) با یکدیگر و همچنین با مدل‌های ANN و ANFIS که هیچ‌گونه پیش‌پردازشی روی پارامترهای ورودی آنها انجام نشده است، پرداخته شد. نتایج نشان داد که هر شش مدل از دقت بالایی برای تخمین تبخیر روزانه برخوردار هستند و از میان شش مدل مزبور، مدل ANFIS-FS با مقدار ضریب تبیین (R^2) ۰/۹۱ و جذر میانگین مربعات خطای (RMSE) ۰/۱۱ چه در مرحله‌ی آموزش و چه در مرحله‌ی آزمون، نسبت به مدل‌های دیگر از دقت بالاتری برخوردار است. اگرچه در این پژوهش برتری مدل‌های پیش‌پردازش‌ناچیز است اما توانایی مشخص نمودن ترتیب اهمیت پارامترهای ورودی، تعیین تعداد تقریباً ۳۷۲۰ داده‌ی معنی‌دار برای آموزش شبکه و یافتن بهترین ترکیب، آزمون گاماتست را می‌تواند به‌عنوان ابزاری مفید برای پیش‌پردازش پارامترهای ورودی برای مدل‌سازی سریع‌تر تبخیر تبدیل کند.

واژه‌های کلیدی: تبخیر، رگرسیون گام به گام، گاماتست، مدل‌های هوشمند.

۱ دانشجوی دکتری آبیاری و زهکشی، دانشگاه شهرکرد، شهرکرد، ایران، ۰۹۱۸۳۵۲۱۱۰۶

mohammad.zamaniyan@gmail.com (نویسنده مسئول)

۲ استادیار گروه مهندسی آب، دانشگاه شهرکرد، شهرکرد، ایران، fatahi2@gmail.com

۳ کارشناس ارشد مهندسی آب، دانشگاه شهرکرد، شهرکرد، ایران، rezafh@gmail.com

۴ کارشناس ارشد مهندسی آب، دانشگاه شهرکرد، شهرکرد، ایران، hoseinpoor_fereshteh@yahoo.com

مقدمه

تبخیر جزء مهمی از چرخه آب‌شناسی است که تخمین آن برای پایش مستمر و مدیریت منابع آب، طراحی سیستم‌های آبیاری و زهکشی و برنامه‌ریزی آبیاری اهمیت فراوانی دارد (Tabari et al., 2010). با این وجود، تخمین تبخیر با استفاده از متغیرهای اقلیمی به دلیل طبیعت غیرخطی پیچیده آن، همواره با چالش روبه‌رو بوده است (Abudu et al., 2011). به‌طور معمول، تخمین تبخیر با استفاده از روش‌های بیلان آبی و انتقال جرم صورت می‌گیرد که یکی از شیوه‌های مستقیم تخمین این ویژگی، استفاده از تشتک می‌باشد (Eslamian et al., 2008). با این وجود، همیشه و در هر مکانی (به‌ویژه مکان‌های دوردست)، امکان استقرار و نگهداری از یک ابزار هواشناسی وجود ندارد. از طرفی، عملکرد تشتک متأثر از محدودیت‌های ابزاری و مسایل عملیاتی از قبیل خطاهای انسانی، خطاهای ابزاری، کدورت آب، استفاده آب داخل تشتک توسط پرندگان و سایر جانوران و نیز مشکلات نگهداری می‌باشد (Doorenbos and Pruitt, 1977). بنابراین، می‌توان تخمین تبخیر را از طریق مدل‌هایی که به‌صورت ذاتی غیرخطی هستند مورد بررسی قرار داد.

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 اندازه‌گیری مقادیر پارامترها، عدم کفایت دانش بشری برای زندگی و تشخیص تمام عوامل مؤثر بر پدیده‌ی مورد نظر و عدم وجود رابطه‌ی هموار^۱ بین مجموعه داده‌های ورودی و خروجی باشد (Jensen et al., 1990). در شرایطی که هیچ‌گونه دانشی از معادلاتی که رفتار سیستم را مشخص سازند در دسترس نباشد، گاماتست به‌عنوان یک ابزار اساسی می‌تواند برای کمک به ساخت مدلی هموار از رفتار سیستمی آن پدیده براساس سری داده‌های اندازه‌گیری‌شده به‌کار رود. هم‌چنین، این 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، مورد استفاده قرار گیرد (Jones et al., 2002).

Remesan و همکاران (۲۰۰۸) با استفاده از روش گاماتست، عوامل مؤثر بر تابش خورشیدی را در حوزه‌ی آبریز

برو^۲ واقع در انگلستان تعیین نمودند. آن‌ها با استفاده از آمار پارامترهای روزانه هواشناسی شامل دما، باران، سرعت باد و تابش برون‌زمینی، و ایجاد مدل‌های مختلف از ترکیب پارامترهای مزبور، بهترین مدل را مدلی تشخیص دادند که دارای تمام پارامترهای بالا 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.

Moghaddamnia و همکاران (۲۰۰۹) برای مدل‌سازی تبخیر روزانه‌ی منطقه‌ی چاه‌نیمه، واقع در استان سیستان و بلوچستان، از شبکه‌های عصبی مصنوعی و سیستم استنتاج تطبیقی عصبی- فازی استفاده نمودند. آن‌ها در پژوهش خود از چهار پارامتر هواشناسی شامل دمای هوا، کمبود فشار بخار اشباع، رطوبت نسبی و سرعت باد در ارتفاع دو متری استفاده کردند و با استفاده از مدل گاماتست، بهترین ترکیب را به‌دست آوردند که شامل پارامترهای کمبود فشار بخار اشباع، رطوبت- نسبی و سرعت باد در ارتفاع دو متری بود. هم‌چنین، با استفاده از آزمون M، تعداد داده‌های مورد نیاز برای مدل‌سازی تبخیر را به‌دست آوردند و در پایان بیان داشتند که نتایج حاصل از شبکه‌ی عصبی برای پیش‌بینی تبخیر، نسبت به مدل‌های تجربی و سیستم استنتاج تطبیقی عصبی- فازی از دقت بالاتری برخوردار است.

Noori و همکاران (۲۰۱۰) غلظت مونوکسید کربن روزانه در ایستگاه قلعهک تهران را به کمک شبکه‌ی عصبی مصنوعی و سیستم استنتاج تطبیقی عصبی- فازی مدل‌سازی نمودند. آن‌ها با استفاده از دو تکنیک گاماتست و رگرسیون گام به گام توانستند تعداد ۱۲ متغیر اولیه‌ی تأثیرگذار بر غلظت مونوکسید کربن روزانه را به‌ترتیب به نه و هفت متغیر برای ورود به شبکه- های عصبی مصنوعی و سیستم استنتاج تطبیقی عصبی- فازی کاهش دهند. در نهایت، بیان داشتند که مدل حاصل از رگرسیون گام به گام و شبکه‌ی عصبی از دقت بیشتری برای مدل‌سازی غلظت مونوکسید کربن روزانه برخوردار است.

Ahmadi و همکاران (۲۰۰۹) روش‌های متعدد گزینش پارامترهای ورودی مانند (گاماتست، نظریه‌ی آنتروپی، AIC^۳ و BIC^۴) بر انتخاب پارامترهای مؤثرتر تابش خورشیدی در حوزه‌ی آبریز برو در انگلستان را مورد بررسی قرار دادند. آن‌ها

²Brue

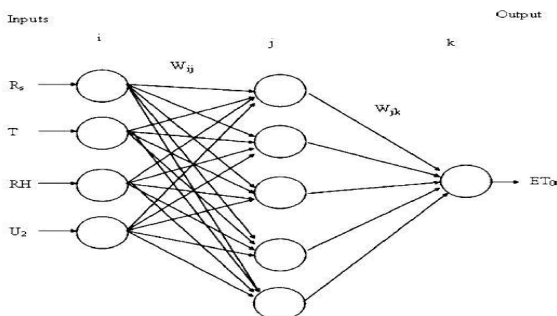
³Akaike's Information Criterion

⁴Bayesian Information Criterion

¹Smooth

تشکیل شده است 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 مسأله مورد بررسی مشخص می‌شوند، حال آن که تعداد نرون‌های موجود در لایه‌های پنهان و همچنین تعداد این لایه‌ها با سعی و خطا در جهت کاهش مقدار خطا توسط طراح مشخص می‌گردند (Moghaddamnia et al., 2009). برخی از کاربردهای ویژه‌ی شبکه‌های عصبی مصنوعی در مدیریت و طراحی منابع آب شامل مدل‌سازی ماهانه، روزانه و ساعتی فرایند روان‌آب بارندگی، پیش‌بینی زمان واقعی مرحله‌ی رودخانه‌ای و دریاچه-ای، پیش‌بینی بارندگی، مدل‌سازی آب‌های زیرزمینی، تشخیص عکس‌العمل هیدرولوژیکی و اکولوژیکی رودخانه نسبت به تغییرات اقلیمی، آنالیزهای خشکسالی و نیز تخمین تبخیر می‌باشند (Kumar et al., 2002; Sudheer et al., 2003; Trajkovic et al., 2003; Keskin and Terzi, 2006; Deswal and Pal, 2008; Rahimi-Khoob, 2009).

طرحی از یک شبکه‌ی عصبی مصنوعی برای تخمین تبخیر و تعرق پتانسیل در شکل (۱) نشان داده شده است (Kisi and Ozturk, 2007).



شکل (۱): طرحی از یک مدل شبکه‌ی عصبی برای تخمین تبخیر و تعرق پتانسیل (R_s تابش خورشیدی، T دما، RH رطوبت نسبی، U_2 سرعت باد در ارتفاع دو متری)

Sudheer و همکاران (۲۰۰۲) مدل شبکه‌ی عصبی را برای پیش‌بینی تبخیر از تشتک کلاس A به کار گرفتند. آن‌ها از متغیرهای مختلف هواشناسی از جمله ساعات آفتابی، رطوبت نسبی، سرعت باد و دمای هوا برای این مدل استفاده کردند.

Keskin و Terzi (۲۰۰۵) برای پیش‌بینی تبخیر از تشتک به این نتیجه رسیدند که مقادیر پیش‌بینی شده با مقادیر واقعی برازش خوبی دارند.

French و همکاران (۱۹۹۲) پیش‌بینی بارندگی در مکان و زمان، Rogers and Dowla (۱۹۹۴) بهینه‌سازی آب‌های زیرزمینی، Shukla و همکاران (۱۹۹۶) طراحی زهکش‌ها در

ضمن مقایسه‌ی روش‌های مختلف بیان داشتند که قابلیت انتخاب بهترین ترکیب از ورودی‌ها، گام‌تست را نسبت به سایر روش‌ها متمایز می‌سازد. هرچند این تکنیک، تعداد داده‌های مورد نیاز برای مدل‌سازی را به‌طور قابل ملاحظه‌ای کم برآورد می‌نماید که در این زمینه، نظریه‌ی آنتروپی دارای قابلیت بهتری است.

در هر حال، پیش‌پردازش داده‌های ورودی برای انتخاب ترکیب مناسب از میان آن‌ها یکی از فرایندهای پیچیده در مدل‌سازی سیستم‌های غیرخطی است. برخلاف وجود مقالات فراوان برای مدل‌سازی پدیده‌ی تبخیر با استفاده از روش‌های هوشمند، همچنان در این باره چند پرسش اساسی وجود دارند، از جمله این که "کدام یک از داده‌های ورودی به مدل از اهمیت بیشتری برخوردار است؟"، "چه ترکیبی از داده‌های ورودی به مدل معرفی شود؟" و یا "چه تعداد داده برای مدل‌سازی این پدیده مورد نیاز است؟".

در اکثر منابع (شایان‌نژاد و همکاران، ۱۳۸۶؛ Rahimi-Khoob, 2009; Traore et al., 2010) برای مدل‌سازی تبخیر با استفاده از روش‌های هوشمند، بهترین ترکیب پارامترهای ورودی با استفاده از روش سعی و خطا بدون به-کارگیری روش‌های پیش‌پردازش انتخاب شده است. بنابراین، فقدان یک روش قاعده‌مند برای یافتن بهترین ترکیب از میان پارامترهای مختلف ورودی احساس می‌شود که این پژوهش به دنبال معرفی روشی برای تعیین بهترین ترکیب از میان پارامترهای مختلف ورودی در برآورد تبخیر با استفاده از روش تشتک و نیز انتخاب بهترین مدل در این ارتباط می‌باشد.

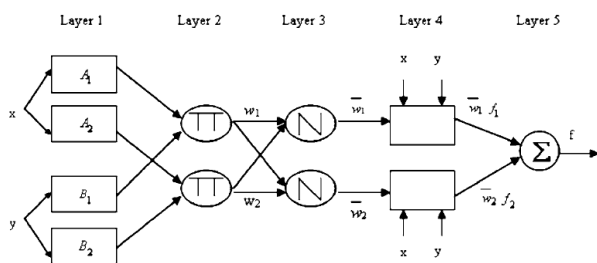
شبکه‌های عصبی مصنوعی^۱ (ANNs)

آغاز نظریه‌ی شبکه‌های عصبی مصنوعی تقریباً به دهه‌ی ۴۰ میلادی برمی‌گردد، زمانی که McCulloch و Rosenblatt معروف و Walter Pitters ریاضی‌دان در سال ۱۹۴۳ آن را پایه-گذاری نمودند (McCulloch and Pitts, 1943). اگرچه ایده-ی شبکه‌ی عصبی مصنوعی، بیشتر از ۶۰ سال پیش ارائه شد ولی نخستین کاربرد عملی شبکه‌های عصبی مصنوعی توسط Rosenblatt با معرفی شبکه‌های پرسپترون چندلایه^۲ (MLP) انجام شد. به‌طور کلی، ساختار شبکه‌ی عصبی پرسپترون چندلایه از سه لایه‌ی ورودی، پنهان و خروجی

^۱Artificial Neural Networks

^۲Multi Layer Perceptron

ساختار مناسب سیستم عصبی- فازی متناسب با داده‌های ورودی، درجه‌ی عضویت، قوانین و توابع درجه‌ی عضویت خروجی انتخاب می‌شود (سبزی‌پرور و همکاران، ۱۳۸۹). شکل (۲) نمایشی از معماری شبکه‌ی عصبی- فازی با دو ورودی، یک خروجی و دو قانون می‌باشد. در لایه‌ی اول (ورودی)، میزان تعلق هر ورودی به بازه‌های مختلف فازی توسط کاربر مشخص می‌شود. با ضرب مقادیر ورودی به هر گره در یکدیگر، وزن قانون‌ها (w_i) در لایه‌ی دوم به دست می‌آید. در لایه‌ی سوم، عمل محاسبه‌ی وزن نسبی قوانین (\bar{w}_i) انجام می‌گیرد. لایه‌ی چهارم، لایه‌ی قوانین است که از انجام عملیات بر روی پیام‌های ورودی به این لایه حاصل می‌شود ($\bar{w}_i f_i$). لایه‌ی آخر، خروجی شبکه (f) می‌باشد که هدف آن حداقل نمودن اختلاف خروجی به دست آمده از شبکه‌ی خروجی واقعی است. در مرحله‌ی آموزش با اصلاح پارامترهای درجه‌ی عضویت براساس میزان خطای قابل قبول، مقادیر ورودی به مقادیر واقعی نزدیک‌تر می‌شوند. روش آموزش اصلی در این سیستم، روش پس‌انتشار خطا^۳ است. در این روش با استفاده از الگوریتم شیب نزولی خطا^۴، مقدار خطا به سمت ورودی‌ها پخش می‌گردد و پارامترها تصحیح می‌شوند (Kisi and Ozturk, 2007).



شکل (۲): شمایی از الگوریتم ANFIS

Kisi (۲۰۰۶) برای تخمین تبخیر روزانه از تشتک، از مدل عصبی- فازی با ورودی‌های دمای هوا، تشعشع خورشیدی، سرعت باد، فشار هوا و رطوبت نسبی استفاده کرد و به این نتیجه رسید که تکنیک عصبی- فازی با داده‌های هواشناسی در دسترس می‌تواند به‌طور موفقیت‌آمیزی در مدل‌سازی تبخیر از تشتک مورد استفاده قرار گیرد. سبزی‌پرور و همکاران (۱۳۸۹) از پارامترهای هواشناسی مانند میانگین دمای حداقل، میانگین دمای حداکثر، رطوبت نسبی، ساعات آفتابی و سرعت باد به-

حالت‌غیرماندگار، Yang و همکاران (۱۹۹۶) مهندسی زهکشی اراضی، Thirumalaian (۱۹۹۸) پیش‌بینی سطح آب در رودخانه‌ها، Jain و همکاران (۱۹۹۹) پیش‌بینی جریان آب ورودی به مخازن را با استفاده از شبکه‌های عصبی مصنوعی ارائه نمودند. Trajkovic و همکاران (۲۰۰۳) برای پیش‌بینی تبخیر و تعرق مرجع از روش شبکه‌های عصبی مصنوعی استفاده نمودند. بدین صورت که برای پیش‌بینی تبخیر و تعرق در روز آینده، از مقادیر تبخیر و تعرق ۱۱ و ۲۳ روز قبل استفاده نمودند. در واقع، لایه‌ی ورودی آن‌ها دارای دو عنصر فرایند بود. نسبت تبخیر و تعرق پیش‌بینی شده به اندازه‌گیری شده، به‌طور متوسط ۰/۹۹۴ بود که کارایی خوب این روش را نشان داد.

سیستم استنتاج تطبیقی عصبی- فازی^۱ (ANFIS)

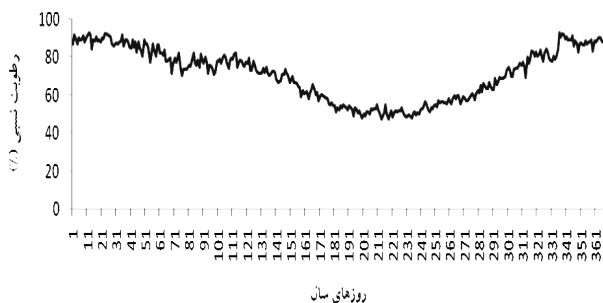
در بین روش‌های نوین مدل‌سازی، سیستم‌های فازی از جایگاه ویژه‌ای برخوردار هستند. توانایی پیاده‌سازی دانش بشری با استفاده از مفاهیم برچسب‌های زبانی و قواعد فازی، غیرخطی بودن و قابلیت سازش‌پذیری این سیستم‌ها و دقت بهتر آن‌ها در مقایسه با سایر روش‌ها در شرایط محدودیت داده‌ها، از جمله مهم‌ترین ویژگی‌های این سیستم‌ها محسوب می‌شود (دزفولی، ۱۳۸۴). نکته‌ی مهم منطق فازی، امکان برقراری ارتباط بین فضای ورودی به فضای خروجی می‌باشد و سازوکار اولیه برای انجام این کار، فهرستی از جمله‌های "اگر- آنگاه"^۲ است که "قانون" نامیده می‌شوند. داشتن روشی که با استفاده از آن بتوان اطلاعات موجود برای ساخت این قواعد را استفاده کرد به‌عنوان ابزاری کارآمد به‌شمار می‌رود. از طرفی، شبکه‌های عصبی مصنوعی به‌دلیل قابلیت‌های آموزش‌پذیری با استفاده از الگوهای مختلف آموزشی می‌توانند ارتباط مناسبی بین متغیرهای ورودی و خروجی ایجاد نمایند. بنابراین، استفاده‌ی ترکیبی از سیستم استنتاج فازی و شبکه‌ی عصبی مصنوعی به‌عنوان ابزاری قدرتمند که قابلیت پیش‌بینی نتایج با استفاده از داده‌های عددی موجود را دارند، تحت عنوان سیستم استنتاج تطبیقی عصبی- فازی معرفی می‌شود. این سیستم از الگوریتم‌های شبکه‌ی عصبی و منطق فازی به منظور طراحی نگاشت غیرخطی بین فضای ورودی و خروجی استفاده می‌کند. سیستم استنتاج تطبیقی عصبی- فازی شبکه‌ای پنج لایه‌ای، متشکل از گره‌ها و کمان‌های اتصال‌دهنده‌ی گره‌ها می‌باشد.

³ Back Propagation

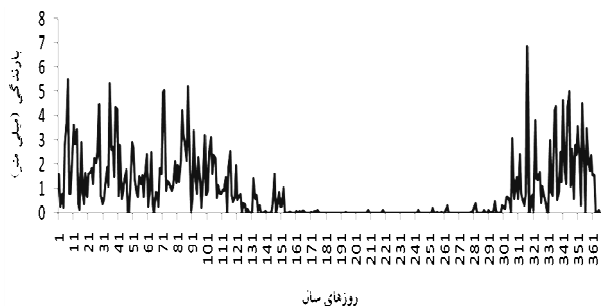
⁴ Gradient Descent Method

¹ Adaptive Neuro- Fuzzy Inference System

² If- Then



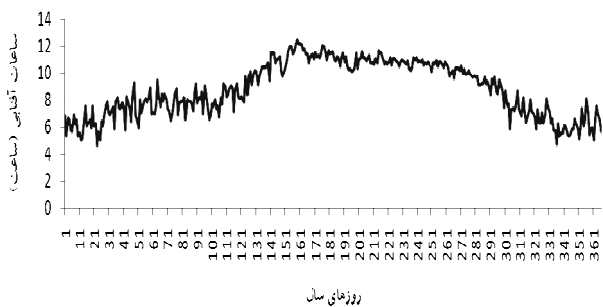
شکل (۴): میانگین تغییرات رطوبت نسبی حداکثر در طول دوره آماری ۱۹۹۵ تا ۲۰۱۰



شکل (۵): میانگین تغییرات بارندگی در طول دوره آماری ۱۹۹۵ تا ۲۰۱۰



شکل (۶): میانگین تغییرات سرعت باد در ارتفاع دو متری در طول دوره آماری ۱۹۹۵ تا ۲۰۱۰



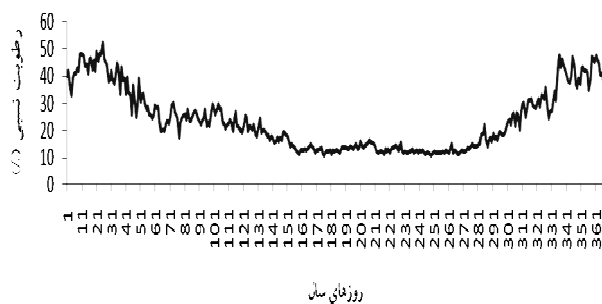
شکل (۷): میانگین تغییرات ساعات آفتابی در طول دوره آماری ۱۹۹۵ تا ۲۰۱۰

عنوان ورودی شبکه‌ی ANFIS و مدل‌های رگرسیونی برای تخمین دمای خاک در سه ایستگاه مختلف رامسر، تهران و زاهدان استفاده نمودند که نتایج آن‌ها نشان داد شبکه‌ی ANFIS در پیش‌بینی دمای خاک، به‌طور متوسط، چهار درصد دقیق‌تر از مدل‌های رگرسیونی عمل می‌کند.

مواد و روش‌ها

مشخصات ایستگاه هواشناسی

برای انجام این پژوهش از داده‌های ایستگاه هواشناسی شهرکرد در استان چهارمحال و بختیاری با مشخصات جغرافیایی طول ۵۰ درجه و ۵۱ دقیقه شرقی و ۳۲ درجه و ۲۰ دقیقه شمالی، ارتفاع ۲۰۴۹ متر از سطح دریا استفاده شده است. داده‌های مورد استفاده شامل دمای حداکثر (T_{max})، دمای حداقل (T_{min})، دمای میانگین (T_{mean})، رطوبت نسبی حداکثر (RH_{max})، رطوبت نسبی حداقل (RH_{min})، رطوبت نسبی میانگین (RH_{mean})، ساعات آفتابی (R_s)، میزان بارندگی (P) و سرعت باد در ارتفاع دو متری (U_2) می‌باشند که به‌صورت روزانه از ابتدای سال ۱۹۹۵ میلادی تا پایان سال ۲۰۱۰ میلادی توسط ایستگاه هواشناسی شهرکرد ثبت و ضبط شده‌اند. خصوصیات آماری داده‌های مورد استفاده در طول دوره آماری مزبور در شکل‌های (۳) تا (۱۰) ارائه شده‌اند. برای پیش‌پردازش داده‌ها نیز از روش‌های رگرسیون گام به گام و گاماتست استفاده شد (Moghaddamnia et al., 2009; Noori et al., 2010).

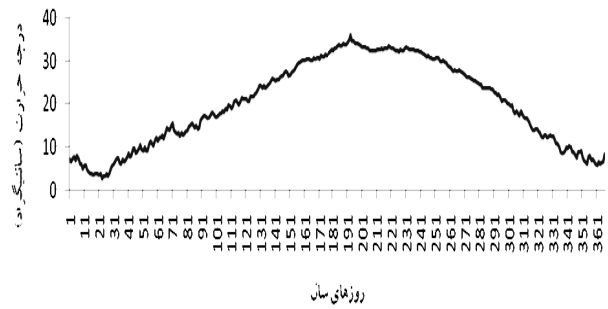


شکل (۳): میانگین تغییرات رطوبت نسبی حداقل در طول دوره آماری ۱۹۹۵ تا ۲۰۱۰

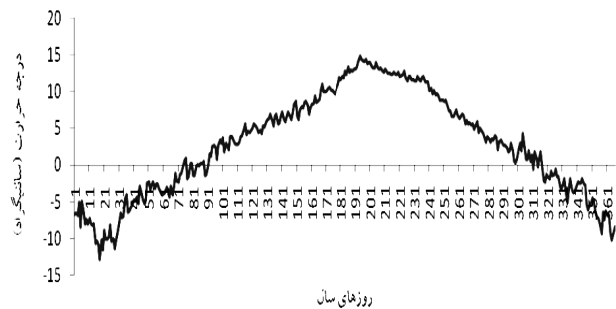
روش‌های نوع گام به گام معروف هستند. روش‌های گام به گام را می‌توان در سه دسته کلی گزینش پیش‌رونده^۱، حذف پس‌رونده^۲ و رگرسیون مرحله‌ای^۳ که ترکیبی از روش‌های پیشین است تقسیم نمود (رضایی و سلطانی، ۱۳۸۲). پژوهشگران مختلفی از جمله Chen و همکاران (۲۰۰۴)، Wang و همکاران (۲۰۰۶) و Khan و همکاران (۲۰۰۷) از روش گزینش پیش‌رونده به‌عنوان ابزاری برای پیش‌بینی و مدل‌سازی پدیده‌ها استفاده کرده‌اند (Noori et al., 2010). در روش رگرسیون گام به گام پیش‌رونده که براساس مدل رگرسیون خطی بنا نهاده شده است، برای قضاوت این که در هر مرحله آیا یک متغیر مستقل جدید باید به مدل افزوده شود یا خیر، یک سطح α انتخاب می‌شود که در این مطالعه، سطح معنی‌داری α برابر ۰/۰۵ انتخاب شده است. 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 هر یک از پارامترهای مستقل، برازش می‌دهند و برای هر مدل رگرسیون ساده‌ی آماری، مقدار آماره-ی 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 انتخاب‌شده از $F_{\alpha}(1, n-2)$ بزرگتر باشد؛ آن‌گاه متغیر مستقل مربوط، به مدل افزوده می‌شود. در غیر این‌صورت، افزودن این متغیر مستقل به مدل چندان مفید نخواهد بود. سرانجام مجموعه‌ای که با افزودن پارامتر مستقل دیگر در سطح α ، افزایش معنی‌داری در آماره-ی F ایجاد نکند، به‌عنوان بهترین ترکیب ورودی برای مدل-سازی انتخاب می‌گردد (بازرگانلاری، ۱۳۸۴).

آزمون گام‌اتست

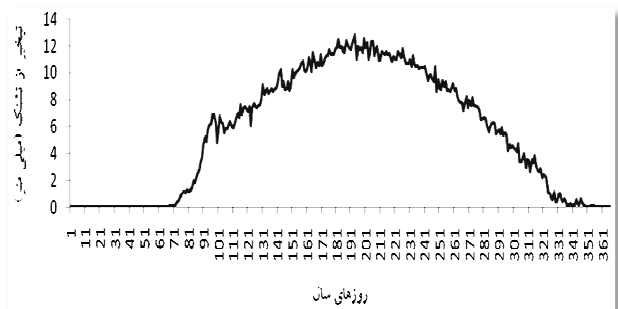
گام‌اتست یک ابزار مدل‌سازی غیرخطی است که به کمک آن می‌توان ترکیب مناسب از پارامترهای ورودی برای مدل-سازی داده‌های خروجی و ایجاد یک مدل هموار را بررسی نمود.



شکل (۸): میانگین تغییرات دمای حداکثر در طول دوره‌ی آماری ۱۹۹۵ تا ۲۰۱۰



شکل (۹): میانگین تغییرات دمای حداقل در طول دوره‌ی آماری ۱۹۹۵ تا ۲۰۱۰



شکل (۱۰): میانگین تغییرات تبخیر روزانه از تشتک در طول دوره‌ی آماری ۱۹۹۵ تا ۲۰۱۰

رگرسیون گام به گام

زمانی که تعداد عوامل ورودی مؤثر (N) بر پدیده‌ای کم باشند، می‌توان با بررسی تمام ترکیبات ممکن و مقایسه‌ی خطا در مدل‌های مختلف رگرسیونی برای پیش‌بینی، آن پدیده را به‌دست آورد. اما وقتی که تعداد پارامترهای ورودی افزایش می‌یابد ارزیابی تمامی رگرسیون‌های ممکن به حجم محاسبات زیادی نیاز دارد. بنابراین، روش‌های مختلفی ایجاد شده‌اند که تنها تعداد کمی از مدل‌های رگرسیونی دارای زیرمجموعه‌ای از متغیرها را در یک زمان بررسی می‌کنند. این روش‌ها عموماً به

¹ Forward Selection

² Backward Elimination

³ Stepwise Selection

مذکور، نشان‌دهنده‌ی مقدار گاماتست است که بیانگر آن بخش از واریانس داده‌های خروجی است که نمی‌تواند به‌وسیله‌ی مدل برآورد گردد. دوم آنکه شیب خط رگرسیون، نشان‌دهنده‌ی پیچیدگی مدلی است که از روی مجموعه داده‌های ورودی و خروجی ساخته می‌شود و این شیب هرچه تندتر باشد نشان‌دهنده‌ی پیچیدگی بیشتر مدل است. یکی دیگر از معیارهای مهم که با استفاده از این آزمون می‌توان به‌دست آورد معیار بدون بعد V_{ratio} است که دارای مقادیری بین بازه‌ی صفر و یک است و هرچه این مقدار به صفر نزدیک‌تر باشد نمایانگر دقت بالای مدل برای یافتن خروجی‌های مطلوب از ورودی‌ها می‌باشد. در واقع، اگر مقدار V_{ratio} از عدد یک کم شود مقدار ضریب تبیین را نشان می‌دهد (Remesan et al., 2008). اگر فرض شود که N پارامتر ورودی بر وقوع پدیده‌ی مؤثر باشند؛ آن‌گاه تعداد $2^N - 1$ ترکیب معنی‌دار از پارامترهای ورودی به‌وجود می‌آیند که برای مدل‌سازی آن با استفاده از مدل‌های هوشمند، بررسی تک‌تک ترکیبات ایجادشده برای یافتن بهترین ترکیب کاری، بسیار وقت‌گیر و خسته‌کننده است. بنابراین، هنگامی که عوامل مؤثر بر پدیده‌ی به‌طور قابل ملاحظه‌ای زیاد باشند با استفاده از گاماتست می‌توان ترتیب میزان اهمیت پارامترهای ورودی و بهترین ترکیب از میان تمام ترکیب‌های ممکن را به‌دست آورد. همچنین، با استفاده از آزمون M می‌توان تعداد داده‌های ورودی را تعیین نمود که مقدار گاما را برای ایجاد مدل هموار به حالت پایدار می‌رسانند (Moghaddamnia et al., 2009).

در این مطالعه، ترکیبات مختلف ورودی مورد آزمون قرار گرفته‌اند تا تأثیر هر پارامتر بر مقدار تبخیر سنجیده شود. برای مشخص نمودن ترتیب اهمیت پارامترهای ورودی، ابتدا آزمون گاماتست برای ترکیبی که همه‌ی پارامترهای ورودی، در آن دخالت دارند انجام شد و سپس در گام بعدی، یکی از پارامترها را به‌دلخواه از مجموعه‌ی اولیه حذف شد و این بار گاماتست با هشت پارامتر باقی‌مانده محاسبه گردید و سپس در گام بعدی، متغیر حذف‌شده به مجموعه پارامترهای اصلی اضافه شد و پارامتر دیگری به‌جای آن حذف می‌گردید. این فرایند برای تمام متغیرها به‌ترتیب انجام گرفت و هر بار، مقدار آماره‌ی گاماتست محاسبه شد. در این فرایند با حذف پارامتر تأثیرگذار، مقدار آماره‌ی گاماتست افزایش می‌یابد و با حذف پارامتری که از اهمیت کمتری برخوردار است مقدار این آماره کاهش خواهد یافت.

همچنین، یک ابزار توسعه‌یافته 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 مشاهده‌شده از پدیده‌ای به‌صورت رابطه‌ی (۱) در اختیار باشد:

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

در صورتی که با توجه به مجموعه‌ی ورودی x_i خروجی y_i به‌دست آید و M ، نشان‌دهنده‌ی تعداد سری مجموعه‌ی مشاهده‌شده از پدیده‌ی مورد بررسی باشد؛ آن‌گاه گاماتست با فرض عدم قطعیت و پذیرش مقدار خطا در به‌دست آوردن خروجی از روی ورودی به‌خاطر پیچیدگی و غیرخطی بودن پدیده‌های مدل‌سازی، این خطا را به‌صورت رابطه‌ی (۲) بین مجموعه‌ی داده‌های ورودی و خروجی با تابع f نشان می‌دهد:

$$y = f(x_1, x_2, \dots, x_m) + r \quad (2)$$

که در آن، f نمایانگر تابع همواری است که برای مدل‌سازی داده‌ها استفاده می‌شود و r نشان‌دهنده‌ی متغیر تصادفی است که برای نمایش خطا به‌کار 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 بالا به مدلی با مشتقات جزئی درجه‌ی اول محدود می‌شود و گاماتست نشان‌دهنده‌ی آن بخش از واریانس داده‌های خروجی است که نمی‌تواند توسط مدل هموار محاسبه گردد. گاماتست براساس $N[i, k]$ می‌باشد که در آن، p بیانگر نزدیکترین همسایگی است. مقادیر این آزمون برای مجموعه داده‌های ورودی و خروجی براساس روابط ۳ و ۴ به‌دست می‌آیند که |...| بیانگر فاصله‌ی اقلیدسی است:

$$\delta_M(k) = \frac{1}{M} \sum_{i=1}^M |x_{N(i,k)} - x_i|^2, 1 \leq k \leq p \quad (3)$$

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

با ایجاد رابطه‌ی رگرسیون خطی بین p مجموعه $(\gamma_M(k), \delta_M(k))$ ، مقدار آماره‌ی گاما برابر عرض از مبدأ خط رگرسیون ایجادشده می‌باشد که معادله‌ی این خط در رابطه‌ی (۵) آورده شده است:

$$\gamma = A\delta + \Gamma \quad (5)$$

از روی خط رگرسیون بیان‌شده در رابطه‌ی بالا می‌توان اطلاعات مفیدی به‌دست آورد: نخست آن‌که عرض از مبدأ خط

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (E_{Modeli} - E_p)^2}{n}} \quad (7)$$

که در آن، \bar{E}_{Modeli} میانگین مقادیر تبخیر روزانه‌ی محاسبه‌شده از مدل‌های مختلف و \bar{E}_p میانگین مقادیر واقعی تبخیر روزانه از تشتک می‌باشند.

در این پژوهش با پیش‌پردازش پارامترهای ورودی توسط روش‌های رگرسیون گام به گام و گاماتست، شش مدل مختلف برای انتخاب بهترین ترکیب از میان پارامترهای ورودی تعیین شدند. اولین گروه مدل‌ها که با استفاده از رگرسیون گام به گام به دست آمد؛ ANN-FS و ANFIS-FS نام دارد. دومین گروه مدل‌ها که بر پایه‌ی پارامترهای مؤثر به دست آمده از گاماتست ساخته شد؛ ANN-GT و ANFIS-GT نام دارد. سومین گروه که بدون انجام هیچ‌گونه پیش‌پردازی بر روی پارامترهای ورودی همه آن‌ها را شامل می‌شود؛ ANN و ANFIS نام دارد. با تعیین پارامترهای مؤثر و ترکیب بهینه در هر روش، به مدل‌سازی تبخیر از تشتک اقدام شد و سپس با معیارهای مختلف آماری سنجش مقدار خطا، به ارزیابی مدل‌های ایجادشده برای تخمین تبخیر روزانه از تشتک پرداخته شد.

نتایج و بحث

آزمون رگرسیون گام به گام

بر اساس نتایج این آزمون در مرحله‌ی نخست، دمای حداکثر مهمترین پارامتری است که برای ورود به مدل انتخاب می‌شود؛ زیرا مطابق با نتایج جدول (۱) که در آن مقادیر ضریب همبستگی ساده (دو به دو) پارامترهای هواشناسی آمده است، پارامتر دمای حداکثر، دارای بیشترین مقدار ضریب همبستگی با تبخیر به مقدار ۰/۸۹ می‌باشد. بنابراین، وقتی به مدل اضافه می‌شود مقدار ضریب تبیین را به ۷۹ درصد می‌رساند (جدول ۲). در گام بعدی، پارامتر حداقل رطوبت نسبی با بیشترین مقدار F نسبت به سایر پارامترهای مستقل باقی‌مانده به مدل اضافه می‌شود که باعث افزایش این ضریب به عدد ۸۰ درصد می‌شود (جدول ۲). 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، میانگین رطوبت نسبی است که ضریب تبیین مدل را به ۸۱ درصد ارتقا می‌دهد

در این پژوهش برای انجام رگرسیون گام به گام از نرم‌افزار آماری SAS و همچنین برای انجام گاماتست از نرم‌افزار Win Gamma که توسط جمعی از پژوهش‌گران گروه کامپیوتر دانشگاه کاردیف برای مدل‌سازی سیستم‌های غیرخطی طراحی گردیده است، استفاده شد. برای مدل‌سازی شبکه‌های عصبی مصنوعی (ANNs) و سیستم استنتاج فازی-عصبی (ANFIS) نیز از نرم‌افزار MATLAB استفاده شده است. این نرم‌افزار امکان طراحی، ساخت، یادگیری و ارزیابی شبکه‌های مذکور را در اختیار می‌دهد و در برگیرنده‌ی شبکه‌های متفاوت با قوانین یادگیری مختلف می‌باشد. در این پژوهش برای مدل‌سازی تبخیر از شبکه‌ی MLP و از توابع محرک کران‌دار تابع سیگموئید و از میان روش‌های مختلف آموزش به روش پس انتشار خطا با الگوریتم لوبز-مارکوارت، با یک لایه‌ی میانی به دلیل همگرایی سریع‌تر در آموزش شبکه و چهار نرون به عنوان بهترین و منطقی‌ترین مدل با کمترین خطا استفاده شده است. علت انتخاب یک لایه‌ی پنهان، این است که چون لایه‌های میانی در ارتباط مستقیم با خروجی نیستند؛ بنابراین تغییرات این لایه، تأثیر چندانی در تعدیل وزن‌ها ایجاد نمی‌کنند (Noori et al., 2009). اساس روش پس انتشار خطا بر پایه‌ی قانون یادگیری اصلاح خطا می‌باشد که از دو مسیر اصلی رفت و برگشت تشکیل می‌شود. در مسیر رفت، بردار ورودی به شبکه اعمال می‌شود و تأثیراتشاز طریق لایه‌های میانی به لایه‌ی خروجی انتشار می‌یابد و بردار خروجی پاسخ واقعی شبکه را تولید می‌نماید. در مدل‌سازی با سیستم استنتاج تطبیقی عصبی-فازی از روش سوگنو^۱ با روش انبوهش حاصلضرب و توابع عضویت گوسی^۲ دوطرفه مرکب (پنج تابع) استفاده شده است.

در این پژوهش به منظور میزان کارایی و عملکرد مدل‌های مختلف، مقادیر تبخیر روزانه‌ی محاسبه‌شده از مدل‌های مختلف (\bar{E}_{Modeli}) با مقادیر واقعی تبخیر روزانه از تشتک (\bar{E}_p) به عنوان مرجع به تعداد n مشاهده، از طریق معیارهای آماری ضریب تبیین (R^2) و جذر میانگین مربعات خطا (RMSE) (رابطه‌ی ۶ و ۷) مقایسه شدند.

$$R \quad (6)$$

¹Sugeno

²Gaussian membership functions

حداکثر رطوبت نسبی و میزان بارندگی نمی‌توانند در سطح آماری ۰/۰۵، سبب ایجاد اختلاف معنی‌دار در مقدار آماره‌ی F شوند.

و به همین ترتیب پارامترهایی که در گام‌های بعدی می‌توانند در سطح α باعث ایجاد معنی‌داری در مقدار آماره F شوند، پارامترهای ساعات آفتابی، دمای حداقل و سرعت باد می‌باشند (جدول ۲) و در نهایت، اضافه نمودن پارامترهای دمای میانگین،

جدول (۱): مقادیر ضریب همبستگی ساده‌ی پارامترهای هواشناسی

پارامتر	T_{max}	T_{min}	T_{mean}	RH_{max}	RH_{min}	RH_{mean}	P	U_2	R_s	E_p
E_p	۰/۸۹	۰/۸۰	۰/۸۸	-۰/۷۳	-۰/۵۹	-۰/۷۲	-۰/۱۹	۰/۱۲	۰/۵۵	۱
R_s	۰/۶۱	۰/۳۴	۰/۵۱	-۰/۵۷	-۰/۷۳	-۰/۷۱	-۰/۴۴	-۰/۳۹	۱	
U_2	۰/۱۲	۰/۲۳	۰/۱۸	-۰/۹۹	-۰/۹۸	-۰/۱۱	۰/۱۱	۱		
P	-۰/۲۳	-۰/۰۴	-۰/۱۵	۰/۲۹	۰/۴۱	۰/۳۸	۱			
RH_{mean}	-۰/۸۲	-۰/۶۲	-۰/۷۶	۰/۹۳	۰/۹۱	۱				
RH_{min}	-۰/۷۳	-۰/۴۷	-۰/۶۴	۰/۶۹	۱					
RH_{max}	-۰/۷۷	-۰/۶۶	-۰/۷۵	۱						
T_{mean}	۰/۹۷	۰/۹۵	۱							
T_{min}	۰/۸۵	۱								
T_{max}	۱									

پارامترهای دمای میانگین، حداکثر رطوبت نسبی و میزان بارش برای مدل‌سازی تبخیر استفاده می‌کند.

بنابراین براساس نتایج جدول (۲)، مدل حاصل از روش رگرسیون گام به گام از تمام پارامترهای ورودی به‌جز

جدول (۲): نتایج آزمون رگرسیون گام به گام پیش‌رونده

R^2	F	P- value	پارامترهای ورودی مدل
۰/۷۸۹	۲۱۸۵۱/۴*	<۰/۰۰۰۱	T_{max}
۰/۷۹۷	۲۲۷/۳۹*	<۰/۰۰۰۱	T_{max}, RH_{min}
۰/۸۰۶	۲۶۸/۴۰*	<۰/۰۰۰۱	$T_{max}, RH_{min}, RH_{mean}$
۰/۸۱۱	۱۴۶/۹۷*	<۰/۰۰۰۱	$T_{max}, RH_{min}, RH_{mean}, R_s$
۰/۸۱۳	۷۵/۱۳*	<۰/۰۰۰۱	$T_{max}, RH_{min}, RH_{mean}, R_s, T_{min}$
۰/۸۱۴	۸/۲۰*	<۰/۰۰۰۱	$T_{max}, RH_{min}, RH_{mean}, R_s, T_{min}, U_2$

* معنی‌داری در سطح آماری ۰/۰۵

بر تبخیر در ایستگاه هواشناسی شهرکرد می‌باشد و پارامترهای دمای حداکثر، رطوبت نسبی حداقل، سرعت باد در ارتفاع دو متری، دمای حداقل، میزان بارندگی، رطوبت نسبی میانگین و رطوبت نسبی حداکثر، به‌ترتیب، در مرتبه‌های بعدی اهمیت قرار دارند.

روش گام‌تست

با استفاده از روش گام‌تست 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 برای ورود به شبکه‌ی عصبی مصنوعی را به‌دست آورد.

نتایج این آزمون برای تبخیر روزانه در جدول (۳) آورده شده‌اند. نتایج این جدول، بیانگر آن است که با حذف پارامتر ساعات آفتابی، آماره‌ی گاما دارای بیشترین مقدار می‌گردد. بنابراین، با توجه به این آزمون، ساعات آفتابی، مؤثرترین پارامتر

جدول (۳): نتایج مقادیر آماره گاما برای ترکیب پارامترهای مختلف ورودی

مقدار گاما	ترکیب پارامترهای ورودی	مقدار گاما	ترکیب پارامترهای ورودی
۰/۰۰۹۱	all	۰/۰۰۹۸	all- T_{max}
۰/۰۰۹۱	all- T_{min}	۰/۰۰۸۹	all- T_{mean}
۰/۰۰۹۰	all- RH_{max}	۰/۰۰۹۶	all- RH_{min}
۰/۰۱	all- R_s	۰/۰۰۹۱	all- RH_{mean}
۰/۰۰۹۱	all- P	۰/۰۰۹۴	all- U_2

تعیین بهترین ترکیب

انتخاب ترکیب مناسب از پارامترهای ورودی، یکی از مهمترین مراحل ساخت و طراحی هرگونه مدل سازی ریاضی و هوشمند است. در این پژوهش با در نظر گرفتن نه پارامتر ورودی مؤثرتر تبخیر از تشتک، تعداد ۵۱۱ ترکیب مختلف ایجاد می شود که برای مدل سازی این پدیده با استفاده از شبکه های عصبی مصنوعی و سیستم استنتاج فازی عصبی بررسی تک تک ترکیبات ایجاد شده برای یافتن بهترین ترکیب، بسیار وقت گیر و خسته کننده است. بنابراین، با

استفاده از الگوریتم ژنتیک با بررسی ترکیب های ممکن ایجاد شده، نتایج حاصل از آن برای چند ترکیب منتخب که دارای کمترین مقدار گاما هستند در جدول (۴) آورده شده اند. براساس نتایج جدول (۴) ترکیبی که شامل پارامترهای ورودی دمای حداکثر و حداقل و سرعت باد در ارتفاع دو متری باشد با کمترین مقدار آماره ی گاما نسبت به سایر ترکیب ها و پایین ترین منحنی مجانب که دارای منحنی مجانب تقریباً همواری است، به عنوان بهترین ترکیب ورودی انتخاب می شود.

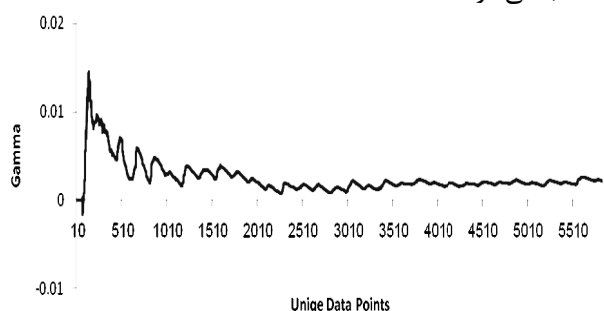
جدول (۴): نتایج مقادیر آماره های گاما، شیب، خطای استاندارد و V_{ratio} برای ترکیب های منتخب

پارامتر	T_{min}, T_{max}, U_2	T_{max}, T_{min}, U_2, P	$T_{max}, T_{min}, U_2, RH_{min}$	$T_{max}, T_{min}, T_{mean}, U_2, P$
گاما	۰/۰۰۵۲	۰/۰۰۵۸	۰/۰۰۶۱	۰/۰۰۶۸
شیب	۰/۶۲۶۸	۰/۵۸۴	۰/۴۶۰	۰/۳۸۶
خطای استاندارد	۰/۰۰۳۴	۰/۰۰۴۱	۰/۰۰۲۹	۰/۰۰۲۳
V_{ratio}	۰/۰۶۸۱	۰/۰۷۶۷	۰/۰۷۹۹	۰/۰۸۹

تعیین تعداد داده های مورد نیاز با استفاده از آزمون M

یکی از چالش های اصلی در مدل سازی پدیده ها، ارزیابی اطلاعات موجود و کفایت داده ها است. به بیان دیگر، چه تعداد الگو از مجموعه الگوهای ورودی باید برای مدل سازی استفاده شود؟ اگر چه به طور کلی با افزودن اطلاعات بیشتر، عملکرد مدل در مرحله ی کالیبراسیون بهبود می یابد، به چه میزان می تواند عملکرد مدل را بهبود بخشد؟ علاوه بر آن، با استفاده از نمودار آزمون M می توان تعداد داده های معنی دار برای ایجاد یک مدل هموار به منظور پیش بینی تبخیر روزانه و چگونگی تغییرات خطای استاندارد و مقدار آماره ی گاما با افزایش تعداد داده ها را نشان داد. با توجه به شکل (۱۱) می توان مشاهده نمود که منحنی های مجانب آماره های گاما و خطای استاندارد، تقریباً در اطراف نقطه ی ۳۷۲۰ به حالت پایدار می رسند که با افزایش تعداد داده ها مجانب های گاما و خطای استاندارد، تغییرات چندانی به خود نمی بیند. بنابراین،

تعداد داده های معنی دار برای آموزش شبکه های عصبی مصنوعی و سیستم استنتاج فازی عصبی، ۳۷۲۰ مجموعه انتخاب می شود.



شکل (۱۱): آزمون M مقادیر گاما برای مدل برگزیده

نتایج پیش پردازش پارامترهای ورودی در مدل های مختلف

براساس نتایج پارامترهای مؤثره دست آمده از روش های رگرسیون گام به گام و گاماتست، بهترین ترکیب از میان

بر پایه‌ی بهترین ترکیب به‌دست آمده از گام‌تست، مدل‌های ANN-GT و ANFIS-GT نیز از ترکیب پارامترهای دمای حداکثر، دمای حداقل و سرعت باد استفاده می‌کنند. مدل‌های ANN و ANFIS نیز بدون انجام هیچ‌گونه پیش‌پردازشی بر روی پارامترهای ورودی، از ترکیب همه‌ی آن‌ها استفاده می‌کنند.

پارامترهای ورودی برای شش مدل مختلف به‌منظور ورود به شبکه‌ی عصبی مصنوعی MLP و سیستم استنتاج فازی عصبی مشخص گردید که نتایج آن برای مدل‌های مختلف در جدول (۵) آمده‌اند. مطابق با نتایج این جدول، مدل‌های ANN-FS و ANFIS-FS برای مدل‌سازی تبخیر از ترکیب پارامترهای دمای حداکثر، حداقل و میانگین رطوبت نسبی، ساعات آفتابی، دمای حداقل و سرعت باد استفاده می‌نمایند.

جدول (۵): مدل‌های برگزیده و پارامترهای ورودی در هر مدل

مدل	ترکیب پارامترهای ورودی
ANFIS-FS و ANN-FS	$T_{max}, RH_{min}, RH_{mean}, R_s, T_{min}, U_2$
ANFIS-GT و ANN-GT	T_{max}, T_{min}, U_2
ANFIS و ANN	$T_{max}, T_{min}, T_{mean}, RH_{max}, RH_{min}, RH_{mean}, P, R_s, U_2$

سنجش خطا در جدول (۶) آورده شده‌اند. نتایج موجود نشان می‌دهند که در هر دو مرحله‌ی آموزش و آزمون، پیش‌پردازش داده‌ها مؤثر بوده است و ضریب تبیین را افزایش داده است. تخمین تبخیر در شبکه‌های عصبی مصنوعی با پیش‌پردازش رگرسیون گام به گام توانسته است نتایج بهتری را به‌دست دهد. براساس نتایج این جدول، هر شش مدل از دقت بالایی برای تخمین تبخیر روزانه برخوردار هستند و از میان آن‌ها، دو مدل ANN-FS و ANFIS-FS با مقادیر ضریب تبیین (R^2) ۰/۸۸ و ۰/۸۸ و جذر میانگین مربعات خطا (RMSE) ۰/۱۰ در مرحله‌ی آزمون، نسبت به مدل‌های دیگر از دقت بالایی برخوردار می‌باشند. همچنین، در مرحله‌ی آزمون، دو مدل ANN-GT و ANFIS-GT دارای ضریب تبیین ۰/۸۵ و جذر میانگین مربعات خطای ۰/۱۱ می‌باشند. برای مرحله‌ی آزمون، مدل ANFIS با ضریب تبیین ۰/۸۴ نسبت به مدل ANN با ضریب تبیین ۰/۸۳، به‌صورت جزئی، نتایج بهتری را ارائه نموده است.

نتایج شبکه‌های عصبی مصنوعی و سیستم استنتاج فازی عصبی

مطابق نتایج حاصل از پیش‌پردازش پارامترهای ورودی در جدول (۵) برای آموزش بخش شبکه‌های عصبی مصنوعی در لایه‌ی ورودی، از ترکیب پارامترهای هواشناسی مختلف استفاده شد. بدین‌منظور از ۵۸۴۴ رکورد ثبت‌شده‌ی ایستگاه هواشناسی شهرکرد، با توجه به آزمون M، تعداد ۳۷۲۰ رکورد برای آموزش و تعداد ۲۱۲۴ رکورد باقی‌مانده برای آزمون شبکه‌های عصبی استفاده شد. برای این‌کار در لایه‌ی پنهان از یک لایه‌ی مخفی با تعداد نرون‌های متفاوتی استفاده شد و تعداد بهینه‌ی آن‌ها برای حداقل نمودن خطا تعیین گردید. روند کار با تعداد نرون‌های کم، آغاز شد و افزودن نرون‌های اضافی تا زمانی ادامه داشت که افزایش نرون‌های بیشتر، تأثیری در بهبود خطا نداشت. بدین‌منظور از ۲ تا ۲۰ نرون استفاده گردید. همچنین، در لایه‌ی پنهان از توابع محرک سیگموئید با روش پس‌انتشار خطا و الگوریتم آموزشی لونیبرگ-مارکواریت مورد استفاده قرار گرفت. نتایج حاصل از مقایسه‌ی شش مدل مذکور با معیارهای آماری مختلف

جدول (۶): مقایسه‌ی مدل‌های مختلف با معیارهای آماری سنجش خطا

مدل	مرحله آموزش		مرحله آزمون	
	RMSE	R^2	RMSE	R^2
ANN-FS	۰/۹۰	۰/۹۰	۰/۰۹	۰/۸۸
ANN-GT	۰/۸۷	۰/۸۷	۰/۱۰	۰/۸۵
ANN	۰/۸۴	۰/۸۴	۰/۱۱	۰/۸۳
ANFIS-FS	۰/۹۱	۰/۹۱	۰/۰۸	۰/۸۸
ANFIS-GT	۰/۸۷	۰/۸۷	۰/۱۰	۰/۸۵
ANFIS	۰/۸۶	۰/۸۶	۰/۱۰	۰/۸۴

غلظت مونوکسید کربن موجود در هوای تهران به این نتیجه رسیدند که پیش‌پردازش داده‌ها با استفاده از روش رگرسیون گام به گام در مقایسه با روش گاماتست با استفاده از مدل-های ANN و ANFIS توانسته عملکرد بهتری داشته باشد.

همچنین، نتایج به‌کارگیری روش‌های پیش‌پردازش مزبور نشان داد که روش FS با مشخص نمودن اهمیت نسبی پیوستن پارامترها به مدل در هر گام و تعیین مدل بهینه‌ی نهایی در سطح معنی‌داری ۰/۰۵، می‌تواند به‌عنوان روشی مؤثر در پاسخ به پرسش‌های مطرح‌شده در این ارتباط مؤثر واقع شود. به‌علاوه، آزمون گاماتست به‌عنوان یک الگوی جدید، توانایی تعیین تعداد الگوهای لازم برای آموزش شبکه برای ایجاد یک مدل هموار با استفاده از آزمون M (در این پژوهش ۳۷۲۰) را دارا می‌باشد و می‌تواند به‌عنوان یک روش کارآمد برای پیش‌پردازش پارامترهای ورودی به‌منظور مدل-سازی تبخیر با استفاده از روش‌های ANN و ANFIS به‌کار گرفته شود.

به‌طور کلی، به‌نظر می‌رسد که ارایه و به‌کارگیری الگوهای نو و تکنیک‌های پیشرفته برای 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.

نتیجه‌گیری

در این پژوهش از روش رگرسیون گام به گام و یک ابزار جدید به نام گاماتست برای پیش‌پردازش پارامترهای ورودی و انتخاب ترکیب بهینه از پارامترهای ورودی برای شبیه‌سازی تبخیر روزانه به کمک شبکه‌های عصبی مصنوعی و سیستم استنتاج فازی-عصبی استفاده شد. آن‌گاه با مرجع قرار دادن تبخیر اندازه‌گیری‌شده از تشتک، عملکرد مدل‌های مختلف برای تخمین تبخیر روزانه مورد ارزیابی و بررسی قرار گرفت. نتایج ارزیابی شش مدل نشان داد که تمامی آن‌ها از دقت بالایی برای شبیه‌سازی تبخیر روزانه برخوردار هستند. هرچند برتری مدل‌های مذکور چندان محسوس نمی‌باشند؛ لیکن مدل‌های ANFIS-FS و ANN-FS دارای برتری نسبت به سایر مدل‌ها بودند.

نتایج پیش‌پردازش پارامترهای ورودی روش‌های رگرسیون گام به گام و گاماتست برای مدل‌سازی نشان داد که هر دو روش توانسته‌اند در بهبود عملکرد مدل‌های شبیه‌سازی تأثیر خوبی داشته باشند و در این بین، روش رگرسیون گام به گام نتایج بهتری در شبیه‌سازی تخمین تبخیر ارایه کرده است. Noori و همکاران (۲۰۱۰) نیز در پیش‌بینی

منابع

- [۱] بازرگان لاری، ع. ۱۳۸۴. رگرسیون خطی کاربردی. چاپ اول، انتشارات مرکز نشر دانشگاه شیراز.
- [۲] دزفولی، ک. ا. ۱۳۸۴. اصول تئوری فازی و کاربردهای آن در مدل‌سازی مسایل مهندسی آب. انتشارات جهاد دانشگاهی، واحد امیرکبیر، چاپ اول.
- [۳] رضایی، ع. و ا. سلطانی. ۱۳۸۲. مقدمه‌ای بر تحلیل رگرسیون کاربردی. دانشگاه صنعتی اصفهان، مرکز نشر.
- [۴] سبزی‌پرور، ع. ا.، ح. زارع ابیانه و م. بیات ورکشی. ۱۳۸۹. مقایسه‌ی یافته‌های مدل شبکه‌ی استنتاج تطبیقی عصبی-فازی با مدل‌های رگرسیونی به‌منظور برآورد دمای خاک در سه اقلیم متفاوت. نشریه آب و خاک (علوم و صنایع کشاورزی)، جلد ۲۴، شماره ۲، صفحه ۲۸۵-۲۷۴.
- [۵] شایان‌نژاد، م. ج. ساداتی‌نژاد و ه. فهمی. ۱۳۸۶. تعیین تبخیر و تعرق بالقوه با استفاده از رگرسیون فازی. مجله تحقیقات منابع آب ایران، شماره ۳، صفحه ۱۹-۹.
- [6] Abudu, S., C. Cui, P. King, J. Moreno and S. Bawazir. 2011. Modeling of daily pan evaporation using partial least squares regression. *Technological Sci.* 54 (1): 163-174.
- [7] Ahmadi, A., D. Han, M. Karamouz and R. Remesan. 2009. Input data selection for solar radiation estimation. *Hydrol. Processes* 23: 2754-2764.
- [8] Deswal, S. and M. Pal. 2008. Artificial neural network based modeling of evaporation losses in reservoirs. *Proc. World Acad. Sci. Eng. Technol.* 29: 279-283.
- [9] Doorenbos, J. and W. O. Pruitt. 1977. Guidelines for prediction of crop water requirements. *FAO Irrig. and Drain. Paper no. 24*, Rome.

- [10] Eslamian, S. S., S. A. Gohari, M. Biabanaki and R. Malekian. 2008. Estimation of monthly pan evaporation using artificial neural networks and support vector machines. *J. Appl. Sci.* 8(19): 3497–3502.
- [11] French, M. N., W. F. Krayewski and R. R. Cuykendall. 1992. Rainfall forecasting in space and time using a neural networks. *J. Hydrol.* 137: 1–37.
- [12] Jain, S. K., A. Das and D. K. Srivastava. 1999. Application of ANN for reservoir inflow prediction and operation. *J. Water Res. Plan. Manage.* 125 (5): 263–271.
- [13] Jensen, M. E., R. D. Burman and R. G. Allen. 1990. Evapotranspiration and irrigation water requirements. ASCE Manual and Report on Engineering Practice No.70. New York.
- [14] Jones, A., D. Evans, S. Margetts and P. Durrant. 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.
- [15] Keskin, M. E. and O. Terzi. 2006. Artificial neural network models of daily pan evaporation. *J. Hydrol. Eng.* 11(1): 65–70.
- [16] Kisi, O. 2006. Daily pan evaporation modeling using a neuro-fuzzy computing technique. *J. Hydrol.* 329: 636–646.
- [17] Kisi, O. and O. Ozturk. 2007. Adaptive Neuro fuzzy Computing Technique for Evapotranspiration Estimation. *ASCE* 133: 4–368.
- [18] Kumar, M., N. S. Raghuwanshi, R. Singh, W. W. Wallender and W. O. Pruitt. 2002. Estimating evapotranspiration using artificial neural networks. *J. Irrig. and Drain. ASCE.* 128 (4): 224–233.
- [19] McCulloch, W. and W. Pitts. 1943. Logical calculus of the ideas immanent in nervous activity. *Bull. Math Biophys.* 5: 33–115.
- [20] Moghaddamnia, A., M. Ghafari-Gousheh, J. Piri, S. Amin and H. Han. 2009. Evaporation estimation using artificial neural networks and adaptive neuro-fuzzy inference system techniques. *J. Advance Water Res.* 32: 88–97.
- [21] Noori, R., A. Karbassi and M. S. Sabahi. 2009. Evaluation of PCA and gamma test techniques on AAN operation for weekly solid waste prediction, *J. Environmental Manage.* 91: 767–771.
- [22] Noori, R., G. Hoshyaripour, K. H. Ashrafi and B. Nadjar-Araabi. 2010. Uncertainty analysis of developed ANN and ANFIS models in prediction of carbon monoxide daily concentration. *J. Atmospheric Environment* 44: 476–482.
- [23] Rahimi-Khoob, A. 2009. Estimating daily pan evaporation using artificial neural network in a semi-arid environment. *Theor. Appl. Climatol.* 98: 101–105.
- [24] Rogers, L. L. and F. U. Dowla. 1994. Optimization of groundwater remediation using artificial neural networks with parallel solute transport modeling. *Water Resour. Res.* 30 (2): 457–481.
- [25] Remesan, R., M. Shamim and D. Han. 2008. Model data selection using gamma test for daily solar radiation estimation. *J. Hydrol. Processes* 22: 4301–4309.
- [26] Shukla, M. B., R. Kok, S. O. Prasher, G. Clark and R. Lacroix. 1996. Use of artificial neural network in transient drainage design. *Trans. ASAE.* 39 (1): 119–124.
- [27] Sudheer, K. P., A. K. Gosain, D. Rangan and S. M. Saheb. 2002. Modeling evaporation using an artificial neural network algorithm. *Hydrol. Process.* 16: 3189–3202.
- [28] Sudheer, K. P., A. K. Gosain and K. S. Ramasastri. 2003. Estimating actual evapotranspiration from limited climate data using neural computing technique. *J. Irrig. Drain. Eng.* 129(3): 214–218.
- [29] Tabari, H., S. Marofi and A. Sabziparvar. 2010. Estimation of daily pan evaporation using artificial neural network and multivariate non-linear regression. *J. Irri. Sci.* 28: 399–406.
- [30] Terzi, O. and M. E. Keskin. 2005. Modelling of daily pan evaporation. *J. Appl. Sci.* 5(2): 368–372.
- [31] Thirumalaian, K. and M. C. Deo. 1998. River stage forecasting using artificial neural network. *J. Hydrol. Eng.* 3 (1): 26–32.
- [32] Trajkovic, S., B. Todorovic and M. Stankovic. 2003. Forecasting of reference evapotranspiration by Artificial Neural Network. *J. Irrig. Drain. Eng.* 129 (6): 454–457.
- [33] Traore, S., Y. M. Wang and T. Kerh. 2010. Artificial neural network for modeling reference evapotranspiration complex process in Sudano-Sahelian zone. *Agri. Water Manage.* 97: 707–714.
- [34] Yang, C. C., S. O. Prasher and R. Lacroix. 1996. Application of artificial neural network to land drainage engineering. *Trans. ASAE.* 39 (2): 525–533.

Input Parameters Preprocessing in Artificial Neural Networks and Adaptive Neuro- Fuzzy Inference System Using Stepwise Regression and Gamma Test Techniques for Estimation of Daily Evaporation

M. Zamaniyan^{1*}, R. Fatahi², A.R. Fatahi³ and F. Hoseinpoor⁴

Abstract

Being a function of different meteorological parameters and their interactions, evaporation is a complex, nonlinear phenomenon. Preprocessing of input parameters to select appropriate combinations is complex when modeling nonlinear systems. Data preprocessing reduces trial and error steps and recognizes most important parameters on noted phenomenon for modeling using 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 and adaptive neuro- fuzzy inference system to estimate daily evaporation (E_p) at Shahrekord meteorological station. To evaluate the effect of input parameters preprocessing in intelligent models using different statistical error criteria, ANN-FS, ANN-GT, ANFIS-FS and ANFIS-GT with preprocessed parameters were compared against each other and also with ANN and ANFIS models without preprocessed parameters. The results showed that all six models have a high degree of precision to estimate daily E_p . ANFIS-FS model represented a determination coefficient (R^2) of 0.91 and root mean square error (RMSE) of 0.11 both of training and test steps. Although the accuracy of models was slightly each other, but the ability of determination of important of input parameters, education and recognition of the best combination of input parameters with 3720 data in this study by gamma test, makes this model a useful tool for fast preprocessing input parameters to model evaporation.

Keywords: intelligent methods, stepwise regression, gamma test, evaporation.

¹ Ph. D. Student of Irrigation and Drainage of Shahrekord University
(*- Corresponding author Email: mohammad.zamaniyan@gmail.com)

² Assistant Professor of Water Engineering Department of Shahrekord University

³ M. Sc of Irrigation and Drainage of Shahrekord University

⁴ M. Sc of Irrigation and Drainage of Shahrekord University