



دانشگاه گوارن و منابع طبیعی گران

مجله پژوهش‌های حفاظت آب و خاک

جلد هجدهم، شماره چهارم، ۱۳۹۰

<http://jwfst.gau.ac.ir>

مدل‌سازی هوشمند تابش خورشیدی با استفاده از آزمون گاما و مقایسه با معادلات تجربی و اسنچی شده در کرمانشاه

محمد قبائی سوق^۱، * ابوالفضل مساعدی^۲ و امیراحمد دهقانی^۳

^۱ دانش‌آموخته کارشناسی ارشد مهندسی منابع آب دانشگاه علوم کشاورزی و منابع طبیعی گران، ^۲ دانشیار دانشکده منابع طبیعی

و محیط‌زیست دانشگاه فردوسی مشهد، ^۳ استادیار گروه مهندسی آب دانشگاه علوم کشاورزی و منابع طبیعی گران

تاریخ دریافت: ۸۹/۱۰/۲۶؛ تاریخ پذیرش: ۹۰/۶/۲۱

چکیده

تابش خورشیدی یکی از مهم‌ترین پارامترهای ورودی در مدل‌های هیدرولوژیکی و رشد گیاهان می‌باشد، ولی در بسیاری از ایستگاه‌ها به دلیل هزینه زیاد تجهیزات اندازه‌گیری تابش خورشیدی و همچنین هزینه نگهداری آن‌ها، اندازه‌گیری مستقیم این پارامتر به‌طور محدود انجام می‌شود. طی دهه‌های گذشته جهت تخمین تابش خورشیدی، معادلات تجربی و مدل‌های غیرخطی مختلفی بر اساس پارامترهای هواشناسی ارائه شده است. در این پژوهش جهت تخمین تابش خورشیدی از آمار روزانه ۵ پارامتر هواشناسی شامل دمای حداقل، دمای حداکثر، رطوبت نسبی، ساعات آفتابی و تابش برون‌زمینی در ایستگاه سینوپتیک کرمانشاه طی سال‌های ۱۹۹۲ الی ۲۰۰۱ استفاده شد. ابتدا با بررسی صحت و کنترل کیفی داده‌های تابش خورشیدی، داده‌های صحیح طی سه مرحله آزمون استخراج گردید. سپس با استفاده از تکنیک پیش‌پردازش آزمون گاما از میان ۳۱ ترکیب مختلف از پارامترهای ورودی، بهترین ترکیب شامل همه پارامترهای ورودی برای مدل‌های غیرخطی رگرسیون محلی (LLR)، شبکه عصبی مصنوعی (ANN) با دو تابع آموزشی گرادیان ادغامی (SCG) و لوببرگ مارکوارت (LM) تعیین شد. آن‌گاه با استفاده از آزمون M تست تعداد ۱۳۰۰ الگو برای آموزش مدل‌های LLR، ANN و اسنچی معادلات تجربی نیز مشخص گردید و از ۸۸۵ الگوی باقیمانده برای

*مسئول مکاتبه: mosaedi@um.ac.ir

ارزیابی مدل‌های ایجاد شده، استفاده شد. در ادامه مدل‌های LLR، ANN (LM)، ANN (SCG) مبتنی بر آزمون گاما و معادلات تجربی واسنجی شده در دوره آزمون، با معیارهای مختلف آماری (ضریب تبیین R^2)، ریشه میانگین مربعات خطا (RMSE) و میانگین قدرمطلق درصد خطا (MAPE)) با هم مقایسه شدند. نتایج نشان داد که با استفاده از معادلات واسنجی شده مبتنی بر ساعات آفتابی می‌توان با دقت نسبتاً بالایی مقادیر R_s روزانه را تخمین زد. علاوه بر این، مدل‌های غیرخطی ANN (LM)، ANN (SCG) و LLR نسبت به معادلات تجربی واسنجی شده دارای مقدار خطای کمتری هستند. ضمن آن‌که مدل ANN (LM) با ضریب تبیین ۰/۹۵۹۹ روی خط (۱:۱) و مقادیر RMSE و MAPE برابر با ۱/۴۲۱۳ مگاژول بر متر مربع در روز و ۶/۷۶۱۶ درصد دارای بیشترین دقت می‌باشد.

واژه‌های کلیدی: تابش خورشیدی، کنترل کیفی داده‌ها، آزمون گاما، رگرسیون خطی محلی، مدل‌سازی هوشمند، کرمانشاه

مقدمه

تابش خورشیدی^۱ (R_s) یکی از مهم‌ترین پارامترهای موثر در توازن حرارتی سیستم جو- زمین می‌باشد که بر بسیاری از فرآیندهای آب و خاک مانند تبخیر، تعرق، ذوب برف و رشد گیاهان اثر می‌گذارد. آگاهی از مقدار دقیق روزانه R_s در بسیاری از زمینه‌ها شامل: هواشناسی، کشاورزی و معماری می‌تواند به متخصصان این علوم کمک فراوانی نماید. تابش خورشیدی را می‌توان به‌طور مستقیم با استفاده از پیرانومتر^۲ اندازه‌گیری نمود. در بسیاری از ایستگاه‌ها به دلیل هزینه زیاد، شرایط نگهداری و واسنجی نمودن، تجهیزات اندازه‌گیری R_s وجود ندارد. از طرفی در بسیاری از ایستگاه‌ها پارامترهای هواشناسی مانند دمای هوا، ساعات آفتابی، رطوبت نسبی و سرعت باد اندازه‌گیری می‌شوند که می‌توان از روی آن‌ها مقدار R_s را با روابط و مدل‌های ریاضی برآورد نمود (یورک‌اغلو و چلیک، ۲۰۰۶).

طی دهه‌های گذشته، محققان زیادی معادلات تجربی مختلفی برای تخمین R_s با استفاده از سایر پارامترهای هواشناسی ارائه نمودند که می‌توان آن‌ها را در سه گروه معادلات مبتنی بر ابرناکی (ردی،

1- Solar radiation

2- Pyranometer

سویت و کاپل، اهنبرگ و بولن)، ساعات آفتابی (انگستروم، پرسکات) و دما (بريستو و کمپل، هارگریوز سامانی، دوناتلی و کمپل، محمود و هوبارد) تقسیم‌بندی نمود. به‌طور کلی معادلات مبتنی بر ساعات آفتابی با ضرایب واسنجی شده نسبت به دو گروه دیگر معادلات مبتنی بر ابرناکی و دما برای تخمین R_s از دقت بالاتری برخوردارند (منجس و همکاران، ۲۰۰۶؛ تانگ و همکاران، ۲۰۱۰).

اگر چه معادلات مبتنی بر ساعات آفتابی از قدرت تخمین بهتری برخوردارند، ولی در برخی از ایستگاه‌ها به دلیل عدم اندازه‌گیری این پارامتر، استفاده از معادلات دمایی ضروری می‌شود. روش موسوم به انگستروم- پرسکات سال‌های زیادی است که به دلیل سادگی و نتایج قابل قبول کاربرد وسیعی در محاسبه R_s دارد (لیو و همکاران، ۲۰۰۹). آکین اوغلو و اسویت (۱۹۹۰) و الاگیب و مانسل (۲۰۰۰)، با توسعه‌ی معادله انگستروم- پرسکات، آن را به صورت فرم‌های درجه دوم و نمایی ارائه نمودند. بریستو و کمپل (۱۹۸۴) یک رابطه غیرخطی را با استفاده از دمای حداقل و حداکثر جهت تخمین R_s ارائه دادند. محققان دیگری نظیر اسوارتمن و اوگانلاد (۱۹۶۷)، چن و همکاران (۲۰۰۴) و عبدالله (۱۹۹۴) با بررسی فرم‌های گوناگون معادلات ارائه شده و دخالت دادن پارامترهای دیگر نظیر رطوبت نسبی سعی در تخمین دقیق‌تر R_s نمودند.

علاوه بر معادلات تجربی، در طی سال‌های اخیر روش‌های جدید پردازشی مبتنی بر معادلات غیرخطی و هوش مصنوعی نیز برای تخمین R_s توسط محققان مختلف به کار گرفته شده‌است. رگرسیون خطی محلی^۱ یکی از تکنیک‌های ناپارامتری است که با تقسیم فضای مدل‌سازی به تعداد زیادی از ناحیه‌های کوچک، خطی‌سازی را در این ناحیه‌ها مورد بررسی قرار می‌دهد. این روش تا حدودی شبیه روش‌های تفاضل محدود و اجزای محدود می‌باشد که برای حل معادلات غیرخطی در معادلات دیفرانسیل کاربرد وسیعی دارند (احمدی و همکاران، ۲۰۰۹). هم‌چنین شبکه‌های عصبی مصنوعی که با الهام از سلول زیستی طراحی شده‌اند، قادرند با ایجاد نگاشت غیرخطی بین متغیرهای ورودی و خروجی، روابط میان آن‌ها را شبیه‌سازی نمایند (تیمویس و همکاران، ۲۰۰۵).

رحمان و مهندس (۲۰۰۸) برای تخمین مقدار R_s منطقه‌ای در عربستان سعودی از ترکیبات مختلف پارامترهای رطوبت نسبی، دمای هوا و شماره روز ژولیوسی به‌عنوان ورودی در شبکه عصبی MLP^۲ استفاده نمودند. آن‌ها در پایان بیان داشتند که شبکه MLP قادر است با استفاده از پارامترهای رطوبت و

1- Local linear regression

2- Multi Layer Perceptron

دما، مقدار R_s این منطقه را با دقت خوبی برآورد نماید. آزاده و همکاران (۲۰۰۹) با استفاده از داده‌های ماهانه پارامترهای مختلف هواشناسی در ۶ ایستگاه سینوپتیک ایران طی سال‌های ۱۹۹۵ الی ۲۰۰۰ مقادیر تابش خورشیدی ماهانه را به کمک شبکه عصبی MLP و معادله انگستروم- پرسکات برآورد و به مقایسه‌ی آن‌ها با هم‌دیگر پرداخته و بیان داشتند که شبکه MLP با میانگین MAPE برابر ۶۷۰ درصد نسبت به معادله انگستروم- پرسکات دارای دقت بالاتری است. بهرنگ و همکاران (۲۰۱۰) با استفاده از پارامترهای مختلف هواشناسی به تخمین R_s روزانه در منطقه دزفول طی سال‌های ۲۰۰۲ تا ۲۰۰۶ به کمک شبکه‌های عصبی MLP، RBF^۱ و معادلات تجربی پرداختند. آن‌ها با تعریف ۶ ترکیب مختلف از ترکیب پارامترهای هواشناسی به‌عنوان ورودی در شبکه‌های MLP، RBF و مقایسه با ۸ معادله تجربی بیان داشتند که بر اساس معیارهای خطا، بهترین شبکه عصبی با MAPE برابر با ۵/۲۱ درصد نسبت به بهترین مدل تجربی با ۱۰/۲۲ درصد از دقت بالاتری برخوردار است. علیزاده و خلیلی (۲۰۰۹) با استفاده از داده‌های اندازه‌گیری شده روزانه R_s در سال‌های ۱۳۷۸ تا ۱۳۸۰ معادله انگستروم- پرسکات را برای ایستگاه سینوپتیک مشهد واسنجی نمودند و هم‌چنین یک معادله رگرسیونی منطقه‌ای را با در نظر گرفتن عوامل مختلف هواشناسی ارائه نمودند. در نهایت با استفاده از داده‌های سال‌های ۱۳۸۱ و ۱۳۸۲ دو مدل ایجاد شده را اعتبارسنجی نمودند. نتایج این محققین نشان داد که دو مدل دارای اختلاف معنی‌داری نبوده و می‌توان با استفاده از ضرایب واسنجی شده معادله انگستروم- پرسکات، R_s را با دقت مناسبی تخمین زد. بیات و میرلطیفی (۲۰۰۹) از شبکه‌های عصبی مصنوعی و ۶ مدل رگرسیونی برای تخمین R_s ایستگاه شیراز در سال‌های ۱۹۸۵ الی ۱۹۹۰ استفاده نمودند. آن‌ها ابتدا از آمار ایستگاه کرج که تا حدودی دارای شرایط اقلیمی مشابه با ایستگاه شیراز می‌باشد، برای آموزش شبکه‌های مختلف عصبی و هم‌چنین واسنجی معادلات تجربی استفاده نمودند. سپس معادلات تجربی واسنجی شده و شبکه‌های آموزش دیده را برای ایستگاه شیراز با آماره‌های مختلف خطا مورد ارزیابی قرار دادند. نتایج آن‌ها نشان داد شبکه عصبی مصنوعی با پارامترهای ورودی تابش برون‌زمینی، ساعات آفتابی واقعی و ساعات آفتابی حداکثر با RMSE برابر با ۲/۳۴ بهترین نتیجه را ارائه می‌دهد.

به‌رغم کاربردهای گسترده مدل‌های ریاضی غیرخطی، مقایسه مدل‌های مختلف به‌علت مراحل زیاد سعی و خطا به امری وقت‌گیر تبدیل شده است. یکی دیگر از چالش‌های اصلی، ارزیابی کفایت داده‌ها برای مدل‌سازی می‌باشد. در این ارتباط این سئوالات می‌تواند مطرح شود: کدام یک از متغیرهای ورودی باید مورد نیاز است؟ چه تعداد داده برای مدل‌سازی نیاز می‌باشد؟ کدام ترکیب از پارامترهای ورودی باید در مدل مورد استفاده قرار گیرد؟ (ریمسان و همکاران، ۲۰۰۸). علاوه بر این، نتایج تحقیقات مختلف نشان می‌دهد که پدیده همپوشانی پارامترهای ورودی یکی از مشکلات اصلی در روش‌های چند متغیره می‌باشد و شناخت متغیرهای وابسته و مستقل برای مدل‌سازی از اهمیت ویژه‌ای برخوردار است.

آزمون گاما^۱ (G.T) یک ابزار مدل‌سازی غیرخطی است که به کمک آن می‌توان حداقل میانگین مربعات خطایی را که توسط مدل هموار می‌تواند برآورد گردد، محاسبه نمود. هم‌چنین با استفاده از آن می‌توان ترکیب مناسب از پارامترهای ورودی برای مدل‌سازی داده‌های خروجی و تعداد داده‌های لازم برای ایجاد یک مدل هموار را تعیین نمود (مقدم‌نیا و همکاران، ۲۰۰۸). آزمون گاما نخستین بار توسط کانسکر در سال ۱۹۹۷ به‌صورت خلاصه معرفی شد و در سال‌های بعد توسط دورانت، جونس و همکاران و ایوانس با جزئیات بیشتر مورد بررسی و استفاده قرار گرفت (ریمسان و همکاران، ۲۰۰۸). در سال‌های اخیر نیز این آزمون برای کاهش مراحل سعی و خطا در مدل‌های هوشمند و غیرخطی توسط محققان مختلف برای مدل‌سازی پدیده‌های گوناگون به‌کار گرفته شد. مقدم‌نیا و همکاران (۲۰۰۸) در برآورد تبخیر، ریمسان و همکاران (۲۰۰۸)، احمدی و همکاران (۲۰۰۹)، مقدم‌نیا و همکاران (۲۰۰۹) در برآورد تابش خورشیدی، نوری و همکاران (۲۰۰۹) در برآورد تولید زباله‌های هفتگی و قبائی سوق و همکاران (۲۰۱۰) در برآورد تبخیر و تعرق از این آزمون استفاده نموده‌اند.

ریمسان و همکاران (۲۰۰۸) با استفاده از آزمون گاما عوامل موثر بر R_s را در حوضه آبریز برو^۲ انگلستان تعیین نمودند. آن‌ها با استفاده از آمار پارامترهای روزانه هواشناسی دما، بارش، سرعت باد و تابش برون‌زمینی و ایجاد مدل‌های مختلف از ترکیب پارامترهای فوق بهترین مدل را مدلی تشخیص دادند که دارای تمام پارامترهای فوق باشد. هم‌چنین با استفاده از آزمون M تست تعداد ۷۶۵ داده را برای آموزش تعیین و سپس با استفاده از دو مدل غیرخطی LLR و ANN به مدل‌سازی R_s پرداختند

1- Gamma Test

2- Brue

که نتایج بیانگر دقت بیشتر مدل LLR و اهمیت بیشتر پارامتر تابش برون‌زمینی بود. احمدی و همکاران (۲۰۰۹) برای انتخاب ترکیب مناسب از ۴ پارامتر موثر بر R_s شامل: دمای دماسنج تر، دمای هوا، تابش برون‌زمینی و فشار هوا، روش‌های پیش‌پردازش GT، نظریه‌ی آنتروپی، AIC^۱ و BIC^۲ را با استفاده از دو مدل غیرخطی LLR و ANN در حوضه آبریز برو انگلستان مورد بررسی قرار دادند و بیان داشتند که روش‌های رایج AIC و BIC برای انتخاب ترکیب مناسب از پارامترهای ورودی با شکست مواجه و با تاکید بر توانمندی نظریه آنتروپی در دیگر پدیده‌های هیدرولوژیکی با تعداد متغیر بیشتر، این نظریه را برای تخمین R_s در این منطقه مناسب ندانسته و به این نتیجه رسیدند که تکنیک GT در این زمینه دارای عملکرد بهتری است. مقدم‌نیا و همکاران (۲۰۰۹) از تکنیک GT برای پیش‌پردازش پارامترهای ورودی موثر بر R_s در حوضه برو استفاده نمودند. آن‌ها با به‌کارگیری نتایج حاصل از روش پیش‌پردازش GT به مقایسه‌ی مدل‌های مختلف LLR، MLP، شبکه‌ی المان^۳، شبکه عصبی خود همبسته برون‌زاد (NNARX)^۴ و سیستم استنتاج تطبیقی عصبی - فازی^۵ (ANFIS) پرداختند و گزارش نمودند که مدل LLR از دقت بالاتری برخوردار است.

با توجه به این که در کشور ما و بسیاری دیگر از مناطق جهان اندازه‌گیری R_s در بسیاری از ایستگاه‌ها انجام نمی‌پذیرد و علاوه بر آن در مواردی ممکن است داده‌های اشتباهی در گزارشات مربوطه مشاهده شود، بنابراین، اهداف این پژوهش به شرح زیر می‌باشد:

۱- به‌کارگیری روشی جدید مبتنی بر ساعات آفتابی علاوه بر روش‌های موجود برای کنترل و بررسی صحت داده‌های اندازه‌گیری شده R_s . ۲- استفاده از تکنیک پیش‌پردازش آزمون گاما جهت انتخاب ترکیب مناسب از پارامترهای ورودی و تعیین داده‌های مورد نیاز برای آموزش مدل‌های غیرخطی LLR، ANN و واسنجی معادلات تجربی در تعیین R_s . ۳- مدل‌سازی R_s روزانه با استفاده از روش‌های LLR، شبکه عصبی مصنوعی با دو تابع آموزشی گرادیان ادغامی (C.G)^۶ و الگوریتم لونبرگ

- 1- Akaikes information criterion
- 2- Bayesian information criterion
- 3- Elman
- 4- Neural network auto-regressive with exogenous
- 5- Adaptive neuro-fuzzy inference system
- 6- Conjugate Gradient

مارکوارت (LM)^۱ و مقایسه با معادلات تجربی واسنجی شده از روی پارامترهای هواشناسی در ایستگاه سینوپتیک کرمانشاه.

مواد و روش‌ها

موقعیت منطقه مورد مطالعه: در این پژوهش از داده‌های ایستگاه سینوپتیک کرمانشاه با مختصات جغرافیایی طول شرقی ۴۷ درجه و ۹ دقیقه، عرض شمالی ۳۴ درجه و ۲۱ دقیقه، ارتفاع ۱۸۱۳ متر از سطح دریا و متوسط بارندگی سالانه ۴۴۵ میلی‌متر استفاده شده است. داده‌های مورد استفاده شامل: دمای حداکثر (T_{max})، دمای حداقل (T_{min})، رطوبت نسبی (RH)، تابش برون‌زمینی (R_a)، ساعات آفتابی (n) و تابش خورشیدی (R_s) به صورت روزانه از ابتدای سال ۱۹۹۲ تا پایان سال ۲۰۰۱ می‌باشند که از سازمان هواشناسی کشور اخذ شده است.

بررسی صحت و کنترل کیفی داده‌ها: در بسیاری از موارد داده‌های هواشناسی به دلیل خطاهای ناشی از عوامل انسانی و تجهیزات اندازه‌گیری از کیفیت بالایی برخوردار نیستند. در این پژوهش به منظور کنترل صحت داده‌های اندازه‌گیری شده R_s ، از روش ارائه شده توسط مرادی (۲۰۰۹) استفاده شده است. در این روش کنترل کیفی داده‌های تابش خورشیدی در سه مرحله مورد بررسی قرار می‌گیرد و تنها داده‌هایی که در هر سه مرحله صحت آن‌ها تأیید شود برای مدل‌سازی و ادامه کار انتخاب می‌شوند. در مرحله‌ی نخست داده‌های R_s با داده‌های تابش برون‌زمینی بر اساس رابطه ۱ مقایسه می‌شوند. در مرحله دوم مقادیر R_s با مقادیر تابش کل رسیده به سطح افقی در یک روز صاف (R_{s0}) مطابق رابطه ۲ مقایسه می‌شوند که در این رابطه مقادیر R_s می‌توانند کمی بیشتر از مقادیر R_{s0} باشد. در مرحله سوم با استفاده از شاخص i_n به ارتباط بین تابش خورشیدی و طول نسبی ساعات آفتابی پرداخته می‌شود که باید بر اساس رابطه ۳ مقادیر آن از مقادیر حد پائین LL بیشتر باشد.

$$0.03 \times R_a \leq R_s < R_a \quad (۱)$$

$$R_s < 1.1 R_{s0} \quad (۲)$$

$$i_n > LL \quad (۳)$$

1- Levenberg-Marquardt

$$i_n = \begin{cases} K_T & \text{if } n < 0.05 \text{ or } K_T < 0.1 \\ \frac{K_T}{N_n} & \text{else} \end{cases} \quad \text{where } N_n = \frac{n}{N}, \quad K_T = \frac{R_s}{R_a} \quad (۴)$$

$$LL = \begin{cases} a \frac{n}{N} & \text{if } \frac{n}{N} \geq c \\ b & \text{else} \end{cases} \quad (۵)$$

در روابط بالا K_T شاخص صاف بودن جو، a مقدار چارک دوم برای مقادیر K_T روزهایی می‌باشد که مقدار طول نسبی ساعات آفتابی از $0.9 > N_n$ مقدار چارک اول برای تمام روزهایی است که مقدار $N_n < 0.1$ باشد و C برابر نسبت b به a می‌باشد (مرادی، ۲۰۰۹).

رگرسیون خطی محلی (LLR): رگرسیون خطی محلی یکی از روش‌های ناپارامتری برای مدل‌سازی است. در این روش همانند روش‌های تفاضل محدود و اجزای محدود، فضای مدل‌سازی به تعداد زیادی از ناحیه‌های کوچک تقسیم شده و خطی‌سازی در این ناحیه‌ها مورد بررسی قرار می‌گیرد. بنابراین LLR یک روش غیرخطی است که سراسر محدوده مدل‌سازی را شامل می‌شود. مدل LLR به سرعت ساخته می‌شود و برای نقاط مورد جستجو اجرا می‌گردد. معمولاً وقتی تعداد داده‌های آموزشی دارای تراکم بالایی باشند، این مدل از دقت بالایی برخوردار است و در مواقعی که تعداد داده‌های آموزش کم باشد دقت آن کاهش می‌یابد. در روش LLR تعیین تعداد همسایگی p تا حد زیادی بستگی به سطح آشفتگی (سطح آشفتگی بالاتر نیاز به p بیشتری دارد)، چگالی نمونه‌ها (نمونه‌های با چگالی کم برای نقاط مورد جستجو به p کمتری نیاز دارند) و پیچیدگی مدل حاکم بر نقاط مورد جستجو دارد (احمدی و همکاران، ۲۰۰۹).

با فرض آن که تعداد نقاط نزدیک‌ترین همسایگی برابر p_{\max} باشد، آن‌گاه معادله خطی ماتریسی زیر را باید حل نمود:

$$\vec{X} \vec{m} = \vec{y} \quad (۶)$$

X ماتریسی با ابعاد $(p_{\max} \times d)$ می‌باشد که p_{\max} تعداد همسایگی و d بعد می‌باشد. y برداری ستونی با طول p_{\max} که بیانگر خروجی می‌باشد و m برداری ستونی است که پارامترهای آن باید برای بدست آوردن نگاهت بهینه به منظور رسیدن به y از X تعیین شوند.

$$\begin{pmatrix} x_{11} & x_{12} & x_{13} & \cdots & x_{1d} \\ x_{21} & x_{22} & x_{23} & \cdots & x_{2d} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x_{p \max, 1} & x_{p \max, 2} & x_{p \max, 3} & \cdots & x_{p \max, d} \end{pmatrix} \times \begin{pmatrix} m_1 \\ m_2 \\ m_3 \\ \vdots \\ m_d \end{pmatrix} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_{p \max} \end{pmatrix}$$

مرتبه ماتریس m برابر تعداد سطرهای مستقلی است که بر یکتایی جواب ماتریس m موثرند. اگر ماتریس X مربعی و غیرواحد باشد، جواب یکتا برای معادله 6 برابر $m = X^{-1}y$ است و اگر ماتریس X غیرمربعی یا واحد باشد، با تصحیح معادله 6 به حل معادله $|Xm - y|^2$ پرداخته می‌شود. پروس (۱۹۵۵) ثابت نمود که جواب یکتای معادله بالا برابر $m = X^\#y$ می‌باشد که $X^\#$ ماتریس شبه‌معکوس است (مقدم‌نیا و همکاران، ۲۰۰۹). در این مطالعه از الگوریتم kd -tree جهت مرتب‌سازی داده‌های آموزش استفاده شده است که شرح این الگوریتم و چگونگی انجام آن در تحقیقات مور (۱۹۹۱) و تسو و همکاران (۲۰۰۲) به‌طور کامل بیان شده است.

شبکه عصبی مصنوعی: از دیدگاه ریاضی شبکه عصبی به‌عنوان یک تقریب‌گر است. توانایی آن در تقریب بین الگوهای یک مساله باعث می‌شود تا بتوان مسائل با پیچیدگی زیاد از قبیل شناسایی الگو، تفکیک الگو، نگاشت غیرخطی، خودسازمان‌دهی و کنترل را انجام داد. به‌طور کلی هر شبکه عصبی مصنوعی از سه لایه ورودی، پنهانی و خروجی تشکیل شده که در هر لایه تعدادی نرون در نظر گرفته می‌شود که به‌وسیله اتصالات به نرون‌های لایه‌های مجاور وصل می‌شوند. تعداد نرون‌های موجود در لایه‌های ورودی و خروجی با توجه به ماهیت مساله مشخص می‌شود، حال آن‌که تعداد نرون‌های موجود در لایه‌های پنهانی و همچنین تعداد این لایه‌ها با سعی و خطا در جهت کاهش مقدار خطا توسط طراح مشخص می‌گردد. محققان مختلفی در بسیاری از کارهای تجربی و تئوری خود نشان دادند که استفاده از یک لایه پنهان برای تخمین هر نوع تابع پیچیده غیرخطی کفایت می‌کند که مهم‌ترین دلیل درستی این نظریه را عدم ارتباط مستقیم لایه‌های پنهانی با خروجی‌های شبکه می‌دانند (نوری و همکاران، ۲۰۰۹). در الگوریتم پس‌انتشار خطا (BP)^۱ توابع آموزشی مختلفی جهت بهینه نمودن وزن‌ها و اریب‌ها به‌کار می‌رود که در این تحقیق از توابع آموزشی گرادیان ادغامی (CG) و الگوریتم لونبرگ مارکوارت (LM) استفاده شده است.

الگوریتم‌های گرادیان ادغامی: الگوریتم معروف BP، وزن‌ها را در جهت منفی گرادیان تعدیل می‌کند. اگر چه تابع عملکرد به سرعت در جهت خلاف گرادیان کاهش می‌یابد، اما این مساله الزامی را برای همگرایی سریع ایجاد نمی‌کند. در الگوریتم‌های CG جستجو در خلال جهت‌های آمیخته و توأم که به طور کلی همگرایی سریع‌تری را نسبت به جهت‌های نزولی ایجاد می‌کنند، صورت می‌گیرد. تمامی الگوریتم‌های گرادیان ادغامی در اولین تکرار، با جستجو در جهت منفی گرادیان شروع می‌شوند.

$$P_0 = -g_0 \quad (7)$$

که P_0 و g_0 بردارهای ورودی و گرادیان اولیه می‌باشند. آن‌گاه جستجو برای تعیین فاصله بهینه برای حرکت در امتداد جهت جستجوی جدید، شروع می‌شود.

$$x_{k+1} = x_k + a_k P_k \quad (8)$$

که x_k و x_{k+1} بردار وزن‌های جاری و اریب‌ها در تکرارهای k و $k+1$ هم‌چنین، a_k و P_k نرخ یادگیری و بردار ورودی جاری می‌باشند. سپس جهت جستجوی بعدی برای ادغام با روش جستجوی قبلی تعیین می‌شود. روش کلی برای تعیین جهت جستجوی جدید، ترکیب جهت نزولی جدید با جهت جستجوی قبلی است.

$$P_k = -g_k + \beta_k P_{k-1} \quad (9)$$

تفاوت الگوریتم‌های مختلف گرادیان ادغامی در شیوه تعیین و محاسبه مقدار β_k است که نسبت نرم حسابی گرادیان جدید به نرم حسابی گرادیان قبلی است (نوری و همکاران، ۲۰۰۹). هر یک از الگوریتم‌های گرادیان ادغامی نیازمند یک رشته جستجو در هر تکرار می‌باشند. الگوریتم کاهشی شیب مقیاس شده (SCG)^۱ برای پرهیز از جستجوهای زمان‌بر طراحی شده است که در این تحقیق از آن استفاده می‌شود.

الگوریتم لونبرگ مارکوارت (LM): در الگوریتم LM همانند دیگر الگوریتم‌های گوسی نیوتون، هدف محاسبه نکردن ماتریس هشین است. تابع عملکرد در این الگوریتم مجموع مربعات خطا می‌باشد، ماتریس هشین و گرادیان به ترتیب از روابط ۱۰ و ۱۱ محاسبه و تقریب زده می‌شوند.

$$H = J^T J \quad (10)$$

$$g = J^T e \quad (11)$$

که J ماتریس ژاکوبین می‌باشد و حاوی مشتقات اول خطاهای شبکه با توجه به وزن‌ها و اریب‌ها می‌باشد و e بردار خطاهای شبکه است. ماتریس ژاکوبین توسط یک تکنیک استاندارد بازگشتی که بسیار ساده‌تر از محاسبه ماتریس هشین است، محاسبه می‌شود. الگوریتم LM از این تقریب برای ماتریس هشین، به صورت زیر استفاده می‌کند.

$$x_{k+1} = x_k - [J^T J + \mu I]^{-1} J^T e \quad (12)$$

چنانچه اسکالر μ صفر باشد، این روش به روش نیوتون تبدیل می‌شود که از تقریب ماتریس هشین استفاده می‌کند. اگر مقدار μ بزرگ باشد، به گرادیان نزولی با اندازه گام کوچک تبدیل می‌شود. μ ، بعد از هر گام موفق کاهش می‌یابد و تنها در صورتی افزایش می‌یابد که تابع عملکرد افزایش یابد. بنابراین تابع عملکرد در هر تکرار از الگوریتم کاهش می‌یابد (نوری و همکاران، ۲۰۰۹).

آزمون گاما: آزمون گاما (GT) 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 قبل از ایجاد مدل بررسی نمود. اگر N پارامتر ورودی بر وقوع پدیده‌ای موثر باشند، تعداد $2^N - 1$ ترکیب معنی‌دار از پارامترهای ورودی به وجود می‌آید که برای مدل‌سازی بررسی تک‌تک ترکیبات ایجاد شده برای یافتن بهترین ترکیب کاری وقت‌گیر است. بنابراین، با استفاده از این تکنیک می‌توان ترتیب اهمیت پارامترهای ورودی، بهترین ترکیب از میان تمام ترکیب‌های ممکن و تعداد داده‌های لازم برای آموزش مدل را بدست آورد. شرح کامل این روش در مقالات مختلف ارائه شده است که از آن جمله می‌توان به مقدم‌نیا و همکاران (۲۰۰۸)، ریمسان و همکاران (۲۰۰۸)، احمدی و همکاران (۲۰۰۹)، مقدم‌نیا و همکاران (۲۰۰۹)، نوری و همکاران (۲۰۰۹) و قبائی سوق و همکاران (۲۰۱۰) اشاره نمود. بنابراین، به منظور رعایت اختصار از شرح معادلات و اثبات آن‌ها خودداری می‌شود. در این روش با تعیین مقدار نزدیک‌ترین همسایگی p ، مقادیر $\gamma_M(k)$ و $\delta_M(k)$ از روی داده‌های خروجی و ورودی محاسبه می‌شود. سپس با ایجاد رابطه‌ی رگرسیون خطی بین P مجموعه‌ی $\gamma_M(k), \delta_M(k)$ ، معادله‌ی آن از رابطه‌ی ۱۳ به دست می‌آید (ریمسان و همکاران، ۲۰۰۸).

$$\gamma = A\delta + \Gamma \quad (13)$$

عرض از مبدأ خط رگرسیون فوق برابر مقدار آزمون گاما و شیب بیشتر آن بیانگر پیچیدگی بیشتر مدل می‌باشد. همچنین مقدار خطای استاندارد (SE)^۱ آن هرچه به عدد صفر نزدیک‌تر باشد با اطمینان بیشتری می‌توان مقدار گاما را نشان‌دهنده‌ی اغتشاش موجود در داده‌ها دانست (مقدم‌نیا و همکاران، ۲۰۰۸). در آزمون M تست با فرض آن‌که M تعداد سری مجموعه مشاهده شده از پدیده مورد بررسی باشد، با استفاده از آزمون گاما می‌توان چگونگی تغییرات منحنی‌های مجانب گاما و خطای استاندارد را برای ترکیب منتخب با افزایش تعداد الگوهای ورودی M بررسی و تعداد الگوهای ورودی که مقدار این آماره‌ها را به حالت پایدار برای ایجاد مدل هموار می‌رساند، تعیین نمود (مقدم‌نیا و همکاران، ۲۰۰۹).

معادلات تجربی: بسیاری از پارامترهای هواشناسی مانند رطوبت نسبی (RH)، درجه ابری بودن (n/N)، دمای هوا و طول ساعات آفتابی بر مقدار تابش خورشیدی دریافتی تاثیر موثرند (لیو و همکاران، ۲۰۰۹). معادلات تجربی ارائه شده بر اساس پارامترهای مورد نیاز به سه دسته معادلات مبتنی بر ابرناکی، دما و ساعات آفتابی تقسیم‌بندی می‌شوند. اگر چه معادلات مبتنی بر ساعات آفتابی از قدرت تخمین بهتری برخوردارند، ولی در برخی از ایستگاه‌ها به دلیل عدم اندازه‌گیری این پارامتر استفاده از معادلات دمایی ضروری می‌شود. معادلات تجربی ارائه شده دارای برخی ضرایب تجربی می‌باشند که تجربه محققان مختلف ثابت کرده است تنها در صورت استفاده از ضرایب واسنجی شده در هر محل، نتایج تخمین R_s به نتایج واقعی نزدیک خواهد بود (دی‌سوزا و همکاران، ۲۰۰۵). در این پژوهش از ۸ معادله تجربی جهت تخمین R_s استفاده شده است که معادلات ریاضی و علائم اختصاری آن‌ها در جدول ۱ ارائه شده است.

جدول ۱- معادلات تجربی مورد استفاده جهت تخمین R_s به همراه معادلات ریاضی و علایم اختصاری

معادلات ریاضی	علامت اختصاری مدل	مدل
$R_s = R_a \left(a + b \left(\frac{n}{N} \right) \right)$	GSR 1	انگستروم - پرسکات
$R_s = R_a \left(a + b * \left(\frac{n}{N} \right) + c * \left(\frac{n}{N} \right)^2 \right)$	GSR 2	آکین اوغلو و اسویت (۱۹۹۰)
$R_s = R_a * a * \exp \left(b * \left(\frac{n}{N} \right) \right)$	GSR 3	الاگیب و مانسل (۲۰۰۰)
$R_s = R_a * a \sqrt{(T_{\max} - T_{\min})}$	GSR 4	آلن (۱۹۹۷)
$R_s = R_a * a \left(1 - \exp \left(-b (T_{\max} - T_{\min})^c \right) \right)$	GSR 5	بریسنتو و کمبل (۱۹۸۴)
$R_s = a + b * \left(\frac{n}{N} \right) + c * RH$	GSR 6	اسوارتمن و اوگاناد
$R_s = R_a \left(a * \ln(T_{\max} - T_{\min}) + b \left(\frac{n}{N} \right)^c + d \right)$	GSR 7	چن و همکاران (۲۰۰۴)
$R_s = R_a \left(a + b \left(\frac{n}{N} \right) + c * RH + d * T_{mean} \right)$	GSR 8	عبداله (۱۹۹۴)

پارامترهای به کار رفته در جدول بالا به شرح زیر می‌باشند:

T_{\min} حداقل دمای روزانه ($^{\circ}C$)، T_{\max} حداکثر دمای روزانه ($^{\circ}C$)، T_{mean} میانگین دمای روزانه ($^{\circ}C$)، RH متوسط رطوبت نسبی روزانه (%)، n ساعات آفتابی اندازه‌گیری شده، N حداکثر ساعات آفتابی محاسبه شده در طول روز، R_s تابش خورشیدی روزانه ($MJ m^{-2} d^{-1}$)، R_a تابش برون‌زمینی ($MJ m^{-2} d^{-1}$) و ضرایب a ، b ، c و d نیز ضرایب تجربی هستند که باید برای هر معادله با توجه به پارامترهای ورودی و خروجی استخراج گردند.

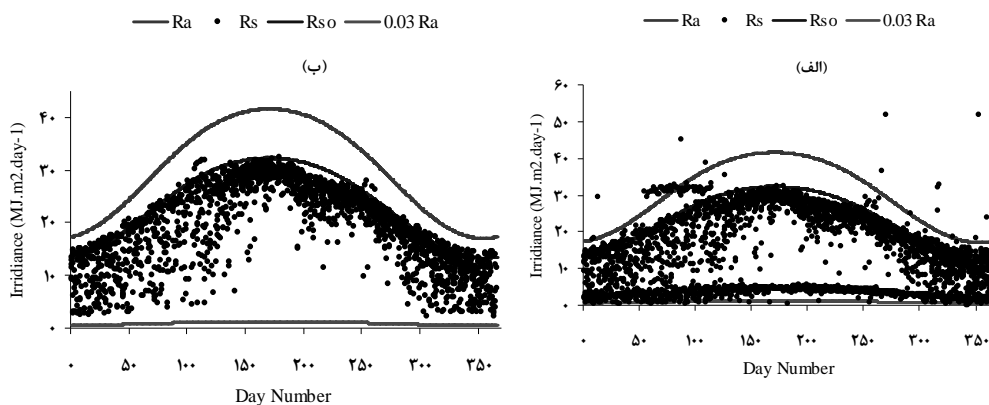
آماره‌های ارزیابی خطا: به منظور سنجش میزان کارایی و ارزیابی عملکرد مدل‌های غیرخطی و معادلات تجربی مقادیر تابش خورشیدی روزانه محاسبه شده از مدل‌ها و معادلات مختلف (R_{SModel}) با مقادیر اندازه‌گیری شده تابش خورشیدی ($R_{SMeasured}$) از طریق معیارهای آماری $RMSE$ ، $MAPE$ و R^2 روی خط ۱:۱ مقایسه گردیدند. بهترین عملکرد مدل بر اساس معیار R^2 یک و براساس معیارهای دیگر صفر است.

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (R_{SModeli} - R_{SMeasuredi})^2}{n}} \quad (14)$$

$$MAPE = \frac{1}{n} \sum_{i=1}^n \frac{|R_{SModeli} - R_{SMeasuredi}|}{(R_{SMeasuredi})} \times 100 \quad (15)$$

نتایج و بحث

نتایج کنترل کیفی داده‌ها و بررسی صحت آن‌ها: پس از به‌کارگیری معادلات ۱ تا ۵ در مورد داده‌های مقادیر تابش برای ایستگاه کرمانشاه، مقادیر a، b و c به‌ترتیب برابر با ۰/۷۱، ۰/۱۳ و ۰/۱۹ به‌دست آمد. در شکل ۱ پراکنش مقادیر تابش در برابر روزهای مختلف سال برای داده‌های اولیه (الف) و تصحیح شده (ب) نشان داده شده‌اند. در شکل الف در ماه‌های میانی سال تعداد داده‌های زیادی بر روی یک خط افقی قرار دارند که در این ماه‌ها با توجه به آن‌که مقدار ساعات آفتابی حداکثر می‌باشد و معمولاً طول روزها بیش از ۱۱ ساعت است، به نظر می‌رسد که داده‌های یادشده اشتباه اندازه‌گیری یا ثبت شده‌اند که این داده‌ها و همچنین داده‌هایی دیگر که از مقدار تابش برون‌زمینی بیشتر هستند با انجام سه سطح آزمون در شکل (ب) پالایش شده‌اند.



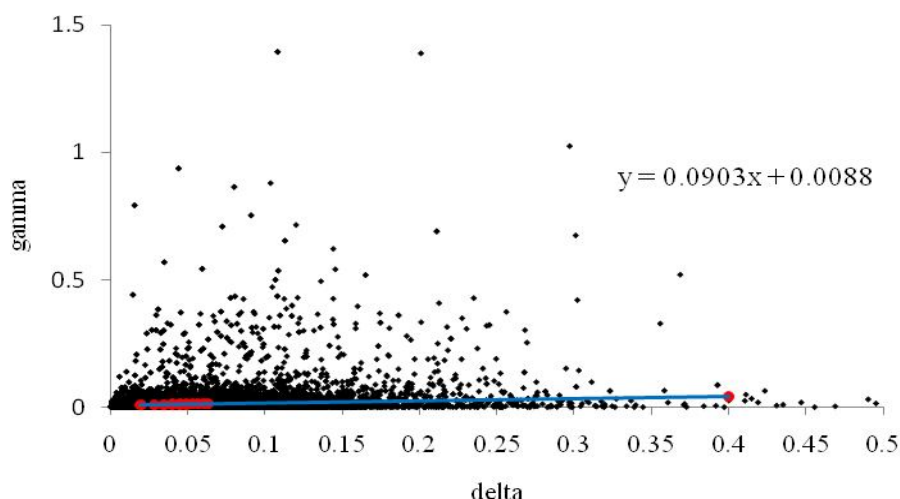
شکل ۱- پراکنش مقادیر تابش در برابر روز الف (داده‌های اولیه)، ب (داده‌های تصحیح شده).

نتایج آزمون گاما: در این تحقیق با در نظر گرفتن پنج پارامتر ورودی موثر بر R_s تعداد ۳۱ ترکیب مختلف ایجاد شد که با استفاده از آزمون گاما با بررسی ترکیبات مختلف ایجاد شده، ترکیبی که دارای کمترین مقدار آماره گاما، گرادیان، SE و V_{Ratio} برای ایجاد یک مدل هموار غیرخطی است، تعیین شد. معیار بدون بعد V_{Ratio} دارای مقادیری بین ۰ و ۱ می‌باشد که هرچه این مقدار به صفر نزدیک‌تر باشد، نمایانگر دقت بالای مدل برای یافتن خروجی‌های مطلوب از ورودی‌هاست. نتایج مقادیر آماره‌های گاما، گرادیان، خطای استاندارد و V_{Ratio} حاصل از این آزمون برای برخی از ترکیبات ایجاد شده در جدول ۲ ارائه شده است که بر اساس آماره‌های فوق، بهترین ترکیب، ترکیبی است که شامل همه پارامترهای موثر ورودی باشد (All).

جدول ۲- نتایج مقادیر آماره‌های گاما، گرادیان، SE و V_{Ratio} برای ترکیبات مختلف پارامترهای ورودی

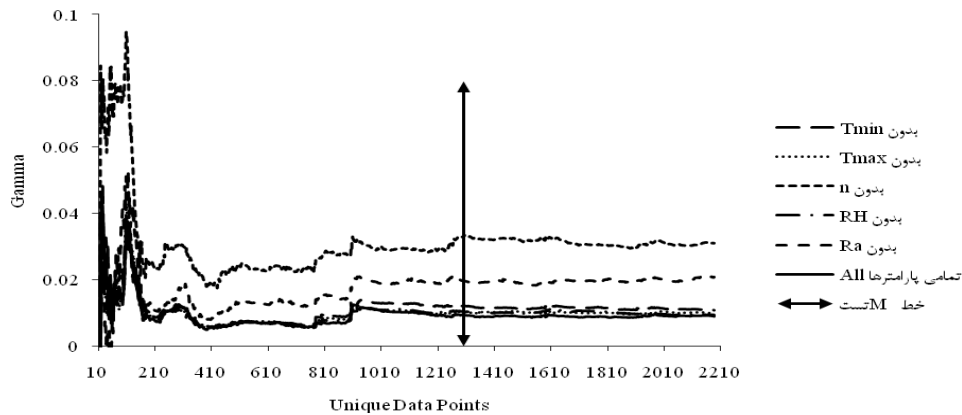
ترکیبات مختلف پارامترهای ورودی						پارامتر
بدون Ra	بدون n	بدون T_{min}	بدون RH	بدون T_{max}	تمامی پارامترها (All)	
۰/۰۲۰۸۳	۰/۰۳۱۰۶	۰/۰۱۰۸۵	۰/۰۰۹۳۴	۰/۰۰۹۹۱	۰/۰۰۸۸۴	گاما
۰/۱۱۵۰۸	۰/۲۶۴۷۶	۰/۰۹۰۵۱	۰/۱۲۲۱۸	۰/۰۹۸۴۳	۰/۰۹۰۳۰	گرادیان
۰/۰۰۰۹۱	۰/۰۰۱۰۳	۰/۰۰۰۸۷	۰/۰۰۱۰۴	۰/۰۰۰۷۲	۰/۰۰۰۵۸	SE
۰/۰۸۳۳۴	۰/۱۲۴۲۲	۰/۰۴۳۴۰	۰/۰۳۷۳۷	۰/۰۳۹۶۳	۰/۰۳۵۳۴	V_{Ratio}

با محاسبه مقادیر دلتا و گاما از روی پارامترهای ورودی و خروجی، پراکنش آن‌ها حول خط رگرسیون برای ترکیب منتخب در شکل ۲ رسم شده است. مطابق این شکل، هرچه تجمع نقاط در حاشیه گوه‌ای شکل در اطراف محور گاما بیشتر باشد، تعداد نقاطی که به ازای ورودی‌های یکسان، خروجی‌های متفاوتی دارند، بیشتر می‌باشد که در شکل ۲ تجمع نقاط در حاشیه گوه‌ای بسیار کم می‌باشد.



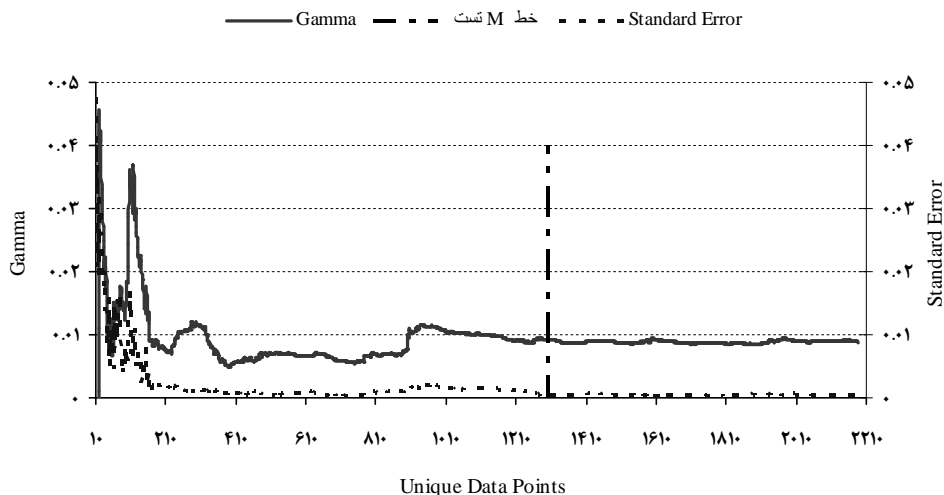
شکل ۲- نمودار پراکنش نقاط در اطراف خط رگرسیون.

برای مشخص نمودن اهمیت پارامترهای ورودی، ابتدا آزمون GT برای ترکیبی که همه‌ی پارامترهای ورودی در آن حضور دارند (All) انجام می‌گردد. سپس در هر مرحله یکی از ۵ پارامتر ورودی به دلخواه از مجموعه اولیه حذف و GT با ۴ پارامتر باقی‌مانده محاسبه شد. این فرآیند برای تمام متغیرها به ترتیب انجام گرفت و هر مرتبه مقدار آماره آزمون گاما محاسبه گردید. نتایج این آزمون برای تابش خورشیدی روزانه در جدول ۲ آورده شده است. با توجه به این جدول با حذف پارامتر ساعات آفتابی، آماره‌ی گاما دارای بیش‌ترین مقدار می‌گردد. بنابراین ساعات آفتابی موثرترین پارامتر بر تابش خورشیدی می‌باشد و پارامترهای تابش برون‌زمینی، دمای حداقل، دمای حداکثر و رطوبت نسبی به ترتیب در مرتبه‌های بعدی اهمیت قرار دارند. منحنی‌های مجانب برای اثر حذف هر یک از پارامترهای ورودی نیز در شکل ۳ ترسیم شده است که مطابق آن منحنی مجانب ترکیب بدون ساعات آفتابی در فاصله بالاتری از منحنی‌های مجانب سایر ترکیبات قرار دارد که نشان‌دهنده اهمیت بیشتر این پارامتر می‌باشد. این یافته با نتایج علیزاده و خلیلی (۲۰۰۹) مطابقت دارد. هم‌چنین سرانو و همکاران (۲۰۰۵) بیان داشتند که تحقیقات نشان داده است که ساعات آفتابی بیشترین تأثیر را روی R_s دارد. در ضمن ترکیبی که شامل همه پارامترهای ورودی است به صورت پائین‌ترین منحنی مجانب ظاهر شده است که بیانگر بهترین ترکیب از میان ۳۱ ترکیب ورودی می‌باشد.



شکل ۳ - منحنی‌های مجانب مقادیر گاما در حالت حذف هریک از پارامترها ورودی

در شکل ۴ می‌توان مشاهده نمود که منحنی‌های مجانب آماره‌های گاما و خطای استاندارد تقریباً در اطراف نقطه ۱۳۰۰ به حالت پایدار می‌رسند که با افزایش تعداد داده‌ها منحنی‌های مجانب آماره‌های گاما و خطای استاندارد تغییرات چندانی به خود نمی‌بینند. بنابراین تعداد الگوهای لازم برای آموزش مدل‌های ANN و LLR و واسنجی معادلات تجربی ۸ گانه برابر با ۱۳۰۰ مجموعه انتخاب می‌شود و از ۸۸۵ الگوی باقی‌مانده برای ارزیابی مدل‌های ایجاد شده با استفاده از آماره‌های خطا استفاده می‌گردد.



شکل ۴ - نمودار آزمون M تست مقادیر گاما و خطای استاندارد برای مدل برگزیده در آزمون گاما

نتایج معادلات تجربی: به‌طور کلی معادلات تجربی براساس همبستگی بین یک و یا چند پارامتر هواشناسی اندازه‌گیری شده و R_s در چندین ایستگاه خاص از یک منطقه ارائه شده‌اند. بنابراین لازم است هنگام کاربرد این معادلات در مناطق جدید آن‌ها را واسنجی و ضرایب محلی را به‌دست آورد. در این تحقیق با استفاده از نتایج آزمون M تست از ۱۳۰۰ الگو داده برای واسنجی معادلات تجربی استفاده شد سپس به‌کمک رگرسیون‌های خطی و غیرخطی موجود در نرم افزار SPSS ضرایب معادلات مذکور محاسبه گردید. ضرایب معادلات تجربی واسنجی شده به‌همراه نتایج آماره‌های مختلف خطا در دوره‌ی واسنجی در جدول ۳ آورده شده است. در ادامه‌ی کار از ۸۸۵ الگوی باقیمانده برای ارزیابی معادلات تجربی واسنجی شده استفاده شد که نتایج آن در جدول ۴ آورده شده است. مطابق نتایج این جدول مشاهده می‌شود که سه معادله‌ی GSR 1، GSR 2 و GSR 3 که از پارامترهای تابش برون‌زمینی و ساعات آفتابی بهره می‌گیرند نسبت به دو معادله‌ی GSR 4 و GSR 5 که مبتنی بر پارامترهای تفاوت دمای حداکثر و حداقل می‌باشند دارای دقتی به مراتب بالاترند. همچنین معادله‌ی GSR 6 که از تابش برون‌زمینی در ترکیب خود استفاده نمی‌نماید دارای کم‌ترین R^2 و بیش‌ترین مقدار RMSE و MAPE خطا می‌باشد. ضمن آن‌که در معادلات GSR 7 و GSR 8 افزودن مؤلفه‌های دما و رطوبت نسبی به معادلات مبتنی بر ساعات آفتابی سبب کاهش مقادیر خطا شده است. معادلات GSR 1 تا GSR 8 در جدول ۱ معرفی شده‌اند. با بررسی معادلات واسنجی شده در دوره‌ی ارزیابی مشاهده شد که معادله‌ی GSR 8 به‌ترتیب با مقادیر R^2 ، RMSE و MAPE برابر با ۰/۹۵۴۴، ۱/۵۵۱۳ و ۷/۵۷۴۰ دارای بیشترین مقدار ضریب تبیین و کمترین مقدار خطا می‌باشد. هرچند این برتری نسبت به معادلات مبتنی بر ساعات آفتابی (GSR 1، GSR 2 و GSR 3) و GSR 7 چندان محسوس نمی‌باشد. بنابراین، در صورت استفاده از معادلات واسنجی شده مبتنی بر ساعات آفتابی، می‌توان با دقت بالایی مقادیر R_s روزانه را تخمین زد که نسبت به معادلات GSR 7 و GSR 8 به پارامترهای کمتری نیاز دارند.

جدول ۳- ضرایب معادلات تجربی واسنجی شده به همراه نتایج مقادیر آماره‌های خطا در دوره‌ی واسنجی

مدل	a	b	c	d	RMSE	MAPE	R ²
GSR 1	۰/۲۸۸	۰/۴۸۹	-	-	۳/۸۳۰۸	۱۰/۸۰۶۳	۰/۹۴۰۰
GSR 2	۰/۲۶۶	۰/۶۲۶	-۰/۱۳۱	-	۲/۱۰۳۴	۱۰/۳۳۶۸	۰/۹۴۰۹
GSR 3	۰/۳۳۴	۰/۸۷۶	-	-	۲/۲۰۳۸	۱۲/۲۴۵۰	۰/۹۳۳۴
GSR 4	۰/۱۵۱	-	-	-	۳/۶۲۱۱	۲۰/۸۴۰۲	۰/۸۰۸۲
GSR 5	۰/۷۴۱	۰/۰۷۲	۱/۱۸۰	-	۳/۵۵۳۴	۱۹/۹۲۴۴	۰/۸۱۰۷
GSR 6	۱۸/۴۴۱	۱۲/۱۶۲	-۰/۱۷۳	-	۴/۲۰۶۶	۲۳/۳۵۵۶	۰/۶۲۰۶
GSR 7	-۰/۰۰۳	۰/۵۱۳	۰/۸۲۸	۰/۲۶۳	۲/۱۰۵۷	۱۰/۳۱۳۲	۰/۹۴۰۸
GSR 8	۰/۴۰۵	۰/۴۹۴	-۰/۰۰۴	-۰/۰۰۱	۲/۰۱۸۹	۱۰/۵۶۹۲	۰/۹۴۱۶

جدول ۴- نتایج ارزیابی معادلات تجربی واسنجی شده در دوره آزمون

آماره	GSR 1	GSR 2	GSR 3	GSR 4	GSR 5	GSR 6	GSR 7	GSR 8
RMSE	۱/۶۲۳۷	۱/۶۳۰۲	۱/۶۷۰۵	۲/۹۷۶۱	۲/۸۵۹۰	۴/۰۸۶۲	۱/۶۲۷۱	۱/۵۵۱۳
MAPE	۷/۸۸۹۳	۷/۶۳۸۸	۸/۷۷۳۱	۱۵/۷۳۲۱	۱۴/۹۴۰۶	۲۱/۶۹۶۴	۷/۵۷۴۰	۷/۵۴۴۸
R ²	۰/۹۵۲۹	۰/۹۵۲۳	۰/۹۴۹۶	۰/۸۴۵۳	۰/۸۴۷۲	۰/۴۸۱۳	۰/۹۵۲۴	۰/۹۵۴۴

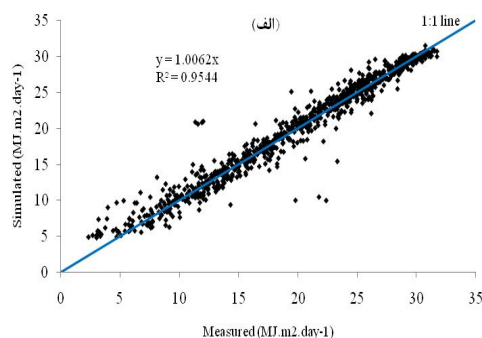
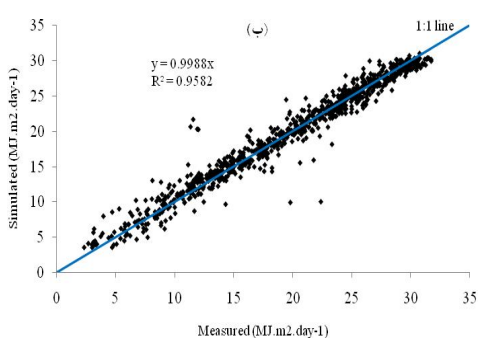
نتایج مدل‌های غیرخطی: با بهره‌گیری از نتایج پیش‌پردازش حاصل از تکنیک آزمون گاما در مرحله‌ی قبل بهترین ترکیب پارامترهای ورودی شامل ۵ پارامتر ورودی اولیه مشخص و مطابق آزمون M تست تعداد ۱۳۰۰ داده برای آموزش مدل‌های شبکه عصبی و رگرسیون خطی محلی تعیین گردید. برای انجام LLR مقدار نزدیک‌ترین همسایگی در جهت کاهش مقدار خطا و افزایش عملکرد با سعی و خطا با نرم‌افزار وین‌گاما^۱ تعیین گردید. بدین منظور مقدار P در محدوده‌ی ۱۰ الی ۴۰ سعی و خطا شد که بهترین مقدار P برابر ۳۶ بدست آمد. برای آموزش مدل‌های ANN از شبکه‌ی عصبی پرسپترون چند لایه با الگوریتم پس‌انتشار خطا و توابع آموزش LM و SCG موجود در نرم‌افزار متلب نسخه ۷/۶ استفاده شد. به منظور طراحی بهینه ساختار شبکه عصبی، از یک لایه میانی با توابع تحریک مختلف تانژانت سیگموئید، لوگ سیگموئید و 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 تاثیری در بهبود خطا نداشته باشند. بدین

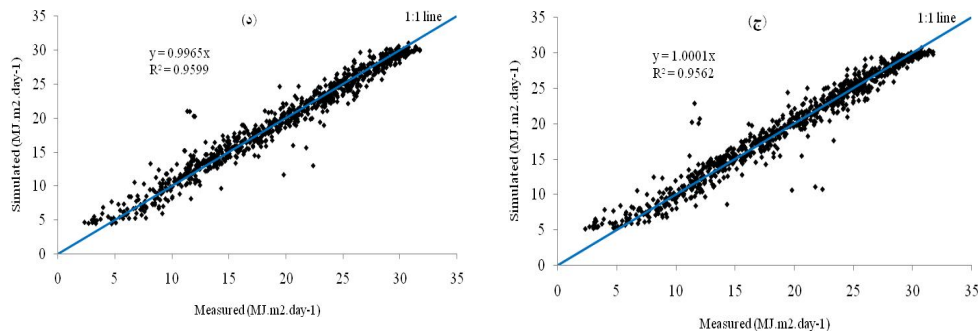
1- <http://users.cs.cf.ac.uk/.../Gamma Software/winGamma/winGamma.htm/>

منظور از ۲ تا ۲۰ نرون در لایه پنهان استفاده گردید. جدول ۵ نتایج مدل‌های شبکه عصبی ANN (LM)، ANN (SCG) و LLR در دوره آزمون را نشان می‌دهد. به منظور مقایسه بهتر نتایج، نمودار پراکنش مقادیر محاسبه شده توسط مدل‌های مختلف در مقابل مقادیر اندازه‌گیری شده تابش خورشیدی حول خطی با شیب ۴۵ درجه در شکل ۵ رسم شده است که هرچه مقادیر محاسباتی و مشاهداتی به هم نزدیک‌تر باشند ابر تشکیل شده در اطراف خط مذکور دارای پهنای کمتر و تراکم بیشتری است. مطابق با جدول ۵ هر سه مدل از دقت بالایی برخوردارند و از میان آن‌ها، مدل ANN (LM) با ضریب تبیین ۰/۹۵۹۹ و مقادیر RMSE و MAPE برابر با ۱/۴۲۱۳ و ۱/۴۲۱۳ مگاژول بر متر مربع و ۶/۷۶۱۶ درصد دارای بیشترین دقت می‌باشد. ضمن آن‌که مدل ANN با تابع آموزشی گرادیان ادغامی نسبت به مدل LLR دارای دقت کمتری است.

جدول ۵- نتایج مدل‌های شبکه عصبی ANN (LM)، ANN (SCG) و LLR در دوره آزمون

مقادیر آماره‌های سنجش میزان خطا			ساختار مدل		مدل
R ²	MAPE	RMSE	نزدیک‌ترین همسایگی		
۰/۹۵۸۳	۶/۹۵۴۵	۱/۴۵۲۴	۳۶		LLR
			تعداد نرون	تابع محرک	
۰/۹۵۹۹	۶/۷۶۱۶	۱/۴۲۱۳	۸	تانژانت سیگموئید	ANN (LM)
۰/۹۵۶۲	۷/۲۷۴۸	۱/۴۹۲۷	۶	لوگ سیگموئید	ANN (SCG)





شکل ۵ - مقایسه عملکرد مدل‌های (الف) GSR 8، (ب) LLR، (ج) ANN (SCG) و (د) ANN (LM) جهت تخمین تابش خورشیدی روزانه بر حسب $(MJ.m^2.d^{-1})$ روی خط ۱:۱ در ایستگاه سینوپتیک کرمانشاه

نتیجه‌گیری کلی

در این تحقیق جهت تخمین مقدار تابش خورشیدی روزانه از سه مدل غیرخطی ((ANN (LM)، ANN (SCG) و LLR) و ۸ معادله تجربی واسنجی‌شده مبتنی بر پارامترهای هواشناسی استفاده شد. استفاده از روش ارائه شده توسط مرادی جهت کنترل صحت داده‌های R_s نشان داد که این روش با منظور نمودن سه سطح آزمون تا حد زیادی می‌تواند به تشخیص داده‌های مشکوک که می‌توانند ناشی از خطاهای انسانی، دستگاه‌های اندازه‌گیری و... باشند، کمک موثری نموده و جایگزین روش‌های کیفی گردد. نتایج حاصل از به‌کارگیری تکنیک پیش‌پردازش آزمون گاما نشان داد که از میان ۳۱ ترکیب مختلف حاصل از پنج پارامتر ورودی، بهترین ترکیب شامل تمام پارامترهای ورودی می‌باشد و از میان پارامترهای ورودی ساعات آفتابی و رطوبت نسبی به ترتیب مهم‌ترین و کم اهمیت‌ترین پارامتر می‌باشند. نتایج ارزیابی معادلات تجربی واسنجی‌شده نشان داد که معادلات مبتنی بر ساعات آفتابی نسبت به معادلات دمایی دارای قدرت تخمین بالاتری هستند و از میان معادلات تجربی، معادله GSR 8 کم‌ترین مقدار خطا را داشت. همچنین مقایسه مدل‌های غیرخطی و معادلات تجربی نشان داد که مدل ANN(LM) با تابع محرک تانژانت سیگموئید دارای بیشترین دقت جهت تخمین تابش خورشیدی می‌باشد. با وجود برتری نتایج سه مدل غیرخطی ANN (LM)، ANN (SCG) و LLR نسبت به معادلات تجربی مورد بررسی، در صورت عدم اندازه‌گیری تمامی پارامترهای مورد نیاز مدل‌های غیرخطی، می‌توان از معادلات واسنجی‌شده نیز به جهت سادگی آن‌ها استفاده نمود.

منابع

1. Abdallah, Y.A.G. 1994. New correlation of global solar radiation with meteorological parameters for Bahrain. *Solar Energy*, 16: 111–120.
2. Ahmadi, A., Han, D., Karamouz, M. and Remesan, R. 2009. Input data selection for solar radiation estimation. *Hydrological Processes*, 23: 2754–2764.
3. Akinoglu, B.G. and Ecevit, A.A. 1990. Further comparison and discussion of sunshine based models to estimate global solar radiation, *Energy*, 15: 865–72.
4. Allen R. 1997. Self calibrating method for estimating solar radiation from air temperature. *Journal of Hydrologic engineering*, 2: 56–67.
5. Alizade, A. and Khalili, N. 2009. Estimation of angstrom coefficient and developing a regression equation for solar radiation estimation (case study: Mashhad). *Journal of Water and Soil*, Vol. 23(1): 229-238.
6. Azadeh, A., Maghsoudi, A. and Sohrabkhani, S. 2009. An integrated artificial neural networks approach for predicting global radiation. *Energy Conversion and Management*, 50: 1497–1505.
7. Bayat, K. and Mirlatifi, M. 2009. Estimation of daily solar radiation using regression models and artificial neural network. *Journal of Agricultural Sciences and Natural Resources*, Vol. 16(3): 270-279.
8. Behrang, M.A., Assareh, E., Ghanbarzadeh, A. and Noghrehabadi, A.R. 2010. The potential of different artificial neural network (ANN) techniques in daily global solar radiation modeling based on meteorological data. *Solar Energy*, 84: 1468–1480.
9. Bristow, K.L., Campbell, G.S., 1984. On the relationship between incoming solar radiation and daily maximum and minimum temperature. *Agric. Forest Meteorol*, 31: 159–166.
10. Chen, R.S., Ersi, K., Yang, J.P., Lu, S.H. and Zhao, W.Z. 2004. Validation of five global radiation models with measured daily data in China. *Energy Convers. Manage*, 45: 1759–1769.
11. De Souza, J.L., Nicacio, R.L. and Lima Moura, M.A. 2005. Global solar radiation measurements in Maceio, Brazil. *Agricultural water Manage*, 30: 1203-1220.
12. Elagib, N. and Mansell, M.G. 2000. New approaches for estimating global solar radiation across Sudan. *Energy Convers Manage*, 41:419–34.
13. Ghabaei sough, M., Mosaedi, A., Hesam, M. and Hezarjaribi, A. 2010. 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. *Journal of Water and Soil*, Vol. 24(3). 610-624.
14. Liu, X., Mei, X., Li, Y., Zhang, Y., Wang, Q., Jensen, J.R. and Porter, J.R. 2009. Calibration of the Angstrom– Prescott coefficients (a, b) under different time scales and their impacts in estimating global solar radiation in the Yellow River basin. *Agricultural and forest meteorology*, 149: 697 – 710.

15. Menges, H.O., Ertekin, C. and Sonmete, M.H. 2006. Evaluation of global solar radiation models for Konya, Turkey. *Energy Conversion and Management*, 47: 3149–3173.
16. Moghaddamnia, A., Ghafari Gousheh, M., Piri, J., Amin, S. and Han, D. 2008. Evaporation estimation using artificial neural networks and adaptive neuro-fuzzy inference system techniques. *Advances in Water Resources*, 32: 88-97.
17. Moghaddamnia, A., Remesan, R., Hassanpour Kashani, M., Mohammadi, M., Han, D. and Piri, J. 2009. Comparison of LLR, MLP, Elman, NNARX and ANFIS Models with a case study in solar radiation estimation. *Journal of Atmospheric and Solar-Terrestrial Physics*, 71: 975–982.
18. Moore, A.W. 1991. An introductory tutorial on k-d Trees. In: Ph.D. thesis: Efficient memory based learning for robot control, Computer Laboratory, University of Cambridge. 209p.
19. Moradi, I. 2009. Quality control of global solar radiation using sunshine duration hours. *Energy*, 34: 1–6.
20. Noori, R., Karbassi, A. and Sabahi, M.S. 2009. Evaluation of PCA and gamma test techniques on AAN operation for weekly solid waste prediction, *Journal of Environmental Management*, 91: 767-771.
21. 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.
22. Rehman, S. and Mohandes, M. 2008. Artificial neural network estimation of global solar radiation using air temperature and relative humidity. *Energy Policy*, 36: 571-576.
23. Serrano, M.L.A., Ruiz, A. Garcia, J.A. Anton M. and Vaquero, J.M. 2005. Solar global radiation and sunshine duration in Extremadura (Spain). *Physica Scripta*, 118: 24-28.
24. Tang, W., Yang, K., He, J., and Qin, J. 2010. Quality control and estimation of global solar radiation in China. *Solar Energy*, 84: 466–475.
26. Tsui, A.P.M., Jones, A.J., and de Oliveira, A.G. 2002. The construction of smooth models using irregular embeddings determined by a gamma test analysis. *Neural Computing & Applications*, 10(4):318–329.
27. Tymvios, F.S., Jacovides, C.P., Michaelides, S.C. and Scouteli, C. 2005. Comparative study of Angstroms and artificial neural networks methodologies in estimating global solar radiation. *Solar Energy*, 78: 752-762.
28. Yorukoglu, M. and Celik, A.N. 2006. A critical review on the estimation of daily global solar radiation from sunshine duration. *Energy Conversion and Management*, 47: 2441-2450.



Gorgan University of Agricultural
Sciences and Natural Resources

J. of Water and Soil Conservation, Vol. 18(4), 2011
<http://jwfst.gau.ac.ir>

Solar radiation data and their intelligent modeling based on gamma test with evaluation of calibrated empirical equations

M. Ghabaei Sough, * A. Mosaedi¹ and A.A. Dehghani

¹Former M.Sc. Student, Dept. of Water Eng. Gorgan University of Agricultural Sciences and Natural Resources, Iran, ²Associate Prof., Faculty of Natural Resources and Environment, Ferdowsi University of Mashhad, Iran, ³Assistance Prof., Dept. of Water Eng. Gorgan University of Agricultural Sciences and Natural Resources, Iran

Received: 2011-1-16; Accepted: 2011-9-12

Abstract

Solar radiation (R_s) is one of the most input parameters in hydrological models and crop growth. Despite its importance for many cases but R_s measurements are not easily available due to the cost, maintenance and calibration requirements of the measuring equipment. Over the past decades, many researchers have developed various equations and none linear models for accurately estimating R_s from meteorological parameters. In this study, daily maximum and minimum air temperature, relative humidity, extraterrestrial radiation and actual sunshine duration values from 1992 to 12001 for Kermanshah synoptic station, were used as inputs. In first stage, the measured R_s data were investigated to control errors and inconsistencies. After quality control test appropriate combination and dataset requiring for training nonlinear model and calibrating empirical equations were determined by using gamma test (GT). Based on GT finding, appropriate combination consists of all input parameters and dataset was needed to training Local Linear Regression (LLR), Artificial Neural Network (ANN) and empirical equations equal to 1300 datasets. The LLR and ANN with two learning methods (Levenberg-Marquardt (LM) and Conjugate Gradient (SCG)) models based on gamma test have been implemented and compared with eight locally calibrated empirical R_s equations. The comparisons have been based on statistical error criteria, using measured daily R_s values. The results indicate that nonlinear model have high accuracy than empirical equation and ANN (LM) with R^2 equal 0.9599 and RMSE and MAPE 1.4213 MJ.m⁻².d⁻¹ and 6.7616 percent respectively has minimum error.

Keywords: Solar radiation, Quality control; Intelligent modeling; Gamma test; Local linear regression; Kermanshah

* Corresponding Author: mosaedi@um.ac.ir