



پیش‌بینی تابش خورشیدی در ایستگاه یزد با بکارگیری مدل رگرسیونی مبتنی بر مؤلفه‌های اصلی (PCR)

سمیه سلطانی گردفرا مرزی^{۱*}

تاریخ دریافت: ۱۴۰۱/۰۴/۲۸

تاریخ پذیرش: ۱۴۰۲/۰۱/۲۴

چکیده

تابش خورشیدی یکی از مهم‌ترین متغیرهای مورد استفاده در مطالعات هیدرولوژی و هواشناسی کشاورزی می‌باشد. در این تحقیق، عملکرد مدل‌های رگرسیونی مؤلفه‌های اصلی (PCR)، در پیش‌بینی مقدار شدت تابش خورشیدی در ایستگاه یزد در دوره‌های ۲۰۱۰ تا ۲۰۱۴ و ۲۰۱۵ تا ۲۰۲۱ بررسی شد. داده‌های مشاهداتی میانگین دما، دمای کمینه، دمای بیشینه، ساعات آفتابی، رطوبت نسبی و تابش خورشیدی و متغیرهای محاسباتی تابش فرازمینی، فاصله نسبی زمین تا خورشید، زاویه میل خورشیدی و حداکثر ساعات آفتابی در مقیاس روزانه به عنوان ورودی روش آنالیز مؤلفه اصلی انتخاب شدند. همبستگی این متغیرها با مقدار شدت تابش خورشیدی نشان از وجود رابطه معنی‌دار مثبت در سطح یک درصد بین مقادیر تابش خورشیدی و تابش فرازمینی، نسبت ساعات آفتابی، دمای میانگین و فاصله نسبی زمین تا خورشید، رابطه مثبت در سطح ۵ درصد با زاویه میل خورشیدی و نیز همبستگی منفی معنی‌دار در سطح ۱ درصد با مقادیر رطوبت نسبی دارد. ضریب همبستگی مؤلفه‌های اول و دوم با تابش خورشیدی به ترتیب ۰/۸۹۳ و ۰/۱۶۸- به دست آمد و در نتیجه مؤلفه اول نسبت به مؤلفه دوم با مقدار تابش خورشیدی همبستگی بیشتری دارد. مقدار بردارهای ویژه دوران شده در مؤلفه اول نشان می‌دهد که دمای کمینه، دمای بیشینه، میانگین دما و رطوبت نسبی با تابش خورشیدی همبستگی بیشتری دارند و همبستگی بقیه عوامل (تابش فرازمینی، زاویه میل خورشیدی و فاصله نسبی زمین تا خورشید و نسبت ساعات آفتابی) در این مؤلفه کم است. بررسی رابطه بین تابش خورشیدی و مقادیر مؤلفه‌های اصلی بدست آمده از PC1 و PC2 با استفاده از رگرسیون خطی چند متغیره نشان داد که اثر متغیرهای هواشناسی بر تابش خورشیدی با ضریب تبیین ۰/۸۹ درصد، معنی‌دار ($p < 0.01$) می‌باشد. معیارهای ارزیابی روش‌های PCR و MLR بیانگر توانایی روش PCR برای برآورد تابش خورشیدی است.

واژه‌های کلیدی: تابش خورشیدی، آنالیز مؤلفه‌های اصلی، عامل تورم واریانس، رگرسیون چند متغیره، یزد

مقدمه

اما علی‌رغم اهمیت اندازه‌گیری آن، وسایل و ابزار مناسب اندازه‌گیری شدت تابش خورشیدی مانند سایر متغیرهای هواشناسی مانند دما و بارش در همه مناطق موجود نیست و بنابراین باید به گونه‌ای آن را برآورد کرد (Liu et al., 2015)؛ Almorox et al., 2005؛ Izimon and Mayer, 2002؛ Mubiru et al., 2007). اغلب اندازه‌گیری‌های انجام شده در زمینه شدت تابش خورشیدی در بسیاری از مناطق به دلیل

اطلاع از مقدار شدت تابش خورشیدی در هر منطقه برای بسیاری از مسائل کاربردی از جمله تخمین تبخیرتغرق و برآورد آب مورد نیاز گیاهان (Almorox et al., 2013)، مدل‌های رشد گیاهان (Liu et al., 2015) و مدل‌سازی‌های هیدرولوژیکی و اقلیمی (Ozgoren et al., 2012) و کاربرد انرژی خورشیدی (Janjai et al., 2011) اهمیت زیادی دارد.

^۱ دانشیار گروه علوم و مهندسی آب، دانشکده کشاورزی و منابع طبیعی

دانشگاه اردکان، اردکان، ایران

(*نویسنده مسئول: ssoltani@ardakan.ac.ir)

نحوه ارجاع مقاله:

سلطانی گردفرا مرزی، س. ۱۴۰۲. پیش‌بینی تابش خورشیدی در ایستگاه یزد با بکارگیری مدل رگرسیونی مبتنی بر مؤلفه‌های اصلی (PCR). نشریه

هواشناسی کشاورزی، ۱۱(۱): ۶-۱۶. DOI: 10.22125/agmj.2023.352446.1140

Soltani-Gerdefamarzi, S. 2023. Prediction of solar radiation intensity in Yazd station by using regression model based on principal components (PCR). Journal of Agricultural Meteorology, 11(1): 6-16. DOI: 10.22125/agmj.2023.352446.1140

دمای هوا، دمای بیشینه هوا و پوشش ابر را برتر و مناسب‌تر از روابط ساده گرسیونی معرفی شده، دانستند. Sabzi-Parvar et al., (2019) عملکرد رگسیون خطی چند متغیره و مدل‌های هوش مصنوعی در تخمین شدت تابش کل خورشیدی را مقایسه کردند. نتایج نشان داد که دقت مدل‌ها در نمونه‌های تصادفی شده تحت شرایط استفاده از همه متغیرهای روزانه شامل میانگین دمای هوا، دمای حداکثر و حداقل، فشار هوا، رطوبت نسبی، بارندگی، دما و رطوبت خاک به عنوان ورودی بهتر است. Ebrahimi et al., (2020) کارایی سامانه استنتاج تطبیقی عصبی-فازی در برآورد شدت تابش کل خورشیدی را مورد بررسی قرار دادند. بر اساس یافته‌های این تحقیق با در نظر گرفتن عملکرد فرآیند مدل‌سازی غیرخطی، مدل ANFIS شدت تابش خورشیدی را نسبت به مدل آنکستروم بهتر برآورد می‌کند. در تحلیل رگسیون، انتخاب متغیرها بسیار مهم است. بدیهی است که خوب بودن مدل رگسیون به متغیرهایی که انتخاب شده است، بستگی دارد. از آنجایی که داده‌های مربوط به همه متغیرهای مؤثر بر شدت تابش خورشیدی در همه ایستگاه‌ها وجود ندارد و یا ثبت نشده است، لذا باید بر روی متغیرهای موجود آنالیز مناسبی انجام داد تا بتوان تأثیر نسبی آنها را بر روی شدت تابش خورشیدی بررسی کرد. بنابراین می‌توان از روش‌های آماری چند متغیره در این خصوص استفاده کرد. روش آنالیز مؤلفه‌های اصلی از متداول‌ترین روش‌های آماری چند متغیره است که روابط مشاهده شده بین چندین متغیر را نشان داده و اهمیت نسبی متغیرهای مؤثر بر یک پدیده را بیان می‌کند. مطالعات انجام شده توسط روش‌های تحلیل عاملی در زمینه‌های هیدرولوژی، هیدرومتئولوژی، منابع آب و محیط زیست بیانگر کارایی این روش می‌باشد. تاکنون از روش آنالیز مؤلفه‌های اصلی برای اهداف متفاوتی نظیر ارزیابی ایستگاه‌های هواشناسی (Seifi et al., 2011)، آلودگی فلزات سنگین در خاک (Soltani-Gerdefaramarzi et al., 2021)، بارندگی و رواناب (Mohan and Arumugam, 1996)، تبخیر تعرق (Sheikholeslami et al., 2014) استفاده شده است. هرچند برای برآورد شدت تابش خورشیدی با استفاده از آنالیز مؤلفه‌های اصلی، مطالعات اندکی وجود دارد. در تحقیق Mohammadi et al., (2018) عملکرد روش تحلیل

مشکلات اقتصادی و فقدان امکانات، حتی در کشورهای در حال توسعه نیز به طور محدود انجام می‌شود (Skeiker, 2006). روش معمول ثبت داده شدت تابش خورشیدی استفاده از دستگاه پیرانومتر است که بدلیل هزینه‌های زیاد نصب و نگهداری در همه مناطق امکانپذیر نیست و مشکلاتی همچون واسنجی آن، تجمع آب و گرد و غبار بر روی سنسور وجود دارد (Rahimikhoob, 2010). حتی در ایستگاه‌هایی هم که تابش را اندازه می‌گیرند، روزهایی وجود دارد که داده‌های شدت تابش ثبت نمی‌شود یا مقادیر غیرواقعی و خارج از بازه مورد انتظار بدلیل نقض دستگاه و یا مشکلات دیگر مشاهده می‌شود (Hunt et al., 1998). به همین دلیل محققین زیادی تشویق به استفاده از مدل‌های تجربی برای برآورد شدت تابش خورشیدی شده‌اند. تحقیقات زیادی به منظور تخمین شدت تابش خورشیدی با استفاده از متغیرهای هواشناسی صورت گرفته و روش‌های زیادی از جمله روابط تجربی و رگرسیونی توسعه یافته است (Sonmete et al., 2010; Menges et al., 2006; Besharat et al., 2013; Yao et al., 2014; Bakirci, 2009). یکی از مشکلات روش‌های برآورد شدت تابش خورشیدی کثرت و متفاوت بودن اطلاعات مورد نیاز است. به‌طوریکه پژوهشگران این روش‌های تخمینی را به چهار دسته تقسیم کرده‌اند. روش‌های وابسته به متغیر دمای هوا، روش‌های وابسته به متغیر ساعات آفتابی، روش‌های بر اساس پوشش ابر و روش‌های ترکیبی وابسته به سایر متغیرهای هواشناسی (Besharat et al., 2013; Chukwujindu, 2017). اولین مدل خطی برای تخمین شدت تابش خورشیدی از داده‌های مدت روشنایی بوسیله Angstrom (1924) معرفی گردید. سپس این مدل توسط پرسکات در سال ۱۹۴۰ با جایگزینی مقادیر تابش فرازمینی با تابش خالص روزانه اصلاح گردید و این مدل به نام مدل آنکستروم-پرسکات بطور وسیع توسط محققین مختلف برای تخمین شدت تابش خورشیدی از داده‌های مختلف هواشناسی استفاده گردید (Trabea and Shaltout, 2000; Li et al., 2011; Chelbi et al., 2015; De Souza et al., 2016; Bakirci, 2015; Medeiros et al., 2017). با توجه به محدودیت مدل‌های رگرسیونی و استفاده آن‌ها برای مناطق مشابه، Okundamiya et al., (2016) روابط رگرسیون چندمتغیره را بر حسب نسبت روشنایی روز، نسبت

می‌باشد. داده‌های ذکر شده از سازمان هواشناسی کشور ایران در دوره زمانی ۲۰۱۰ تا ۲۰۲۱ به صورت روزانه از ایستگاه یزد اخذ گردید. شاخص‌های آماری داده‌های مورد استفاده در جدول ۱ نشان داده شده است. طبق داده‌های جدول کمترین و بیشترین مقدار شدت تابش خورشیدی در ایستگاه یزد ۴/۰۵ و ۳۰/۶۶ و میانگین آن ۱۹/۳۵ مگاژول بر مترمربع بر روز است. کمترین و بیشترین میزان تابش فرازمینی ۳۱/۶۱ و ۳۳/۰۳ و میانگین آن ۳۲/۳۲ مگاژول بر مترمربع بر روز بدست آمد. میانگین نسبت ساعات آفتابی به حداکثر ساعات آفتابی در این ایستگاه ۰/۷۵۸ مشاهده گردید. با توجه به اقلیم خشک ایستگاه یزد، متوسط رطوبت نسبی و دمای منطقه به ترتیب ۲۷/۳۲ درصد و ۲۸ درجه سلسیوس گزارش می‌شود. بر اساس مشکلات ذکر شده در بخش مقدمه، داده‌های تابش خورشیدی ثبت شده توسط پیرانومتر شاید خارج از دامنه مورد انتظار باشد. بنابراین داده‌های تابش خورشیدی لازم است کنترل شوند. به منظور تعیین مقادیر نادرست تابش خورشیدی، شاخص شفافیت روزانه (K_t) محاسبه گردید و مقادیر خارج از بازه $0 < K_t < 1$ حذف شدند (Jiang, Mohammadi et al., 2015; 2009). لازم به ذکر است که این شاخص، نسبت تابش خورشیدی روزانه مشاهده شده به تابش خورشیدی فرازمینی روزانه می‌باشد ($K_t = R_s/R_a$). از آمار سال‌های ۲۰۱۰ تا ۲۰۱۴ برای واسنجی معادله‌ها و از آمار ۲۰۱۵ تا ۲۰۲۱ برای ارزیابی نتایج استفاده گردید. هم‌چنین مقادیر تابش فرازمینی و حداکثر ساعات روشنایی روزانه که وابسته به عرض جغرافیایی محل و شماره روز سال بر مبنای تقویم میلادی می‌باشند، با معادله‌های ۱ تا ۵ محاسبه شدند (Duffie and Beckman, 1991).

$$R_a = \frac{24 \times 3600}{\pi} \cdot I_{gs} \cdot d_r \times \quad (1)$$

$$\left[\cos \varphi \cos \delta \sin W_s + \frac{\pi}{180} \sin \varphi \sin \delta W_s \right] \quad (1)$$

$$d_r = 1 + 0.033 \times \cos\left(\frac{360J}{365}\right) \quad (2)$$

$$\delta = 23.45 \times \sin\left(\frac{360 \times (J + 284)}{365}\right) \quad (3)$$

$$W_s = \cos^{-1}(-\tan \delta \tan \varphi) \quad (4)$$

$$N = \frac{2}{15} \times W_s \quad (5)$$

مؤلفه اصلی و تئوری آنتروپی برای تعیین ورودی مدل‌های شبکه عصبی مصنوعی پرسپترون چندلایه، شبکه عصبی مصنوعی تابع پایه شعاعی، ماشین بردار پشتیبان و برنامه‌ریزی ژنتیک در برآورد شدت تابش خورشیدی در دو ایستگاه همدید کرمان و مشهد مورد بررسی قرار گرفت و هر دو روش برای تعیین ورودی مدل‌های تخمین‌گر به منظور تخمین تابش خورشیدی روش مناسبی تشخیص داده شدند. (Abdelhafidi et al., 2021) با استفاده از رگرسیون خطی چند متغیره گام به گام و آنالیز مؤلفه‌های اصلی، شدت تابش خورشیدی تحت دو سناریوی مختلف در الجزایر را تخمین زدند و روابط استخراج شده را برای مناطق مشابه با اقلیم منطقه مورد مطالعه مناسب دانستند. این پژوهش به بررسی تأثیر نسبی متغیرهای مختلف شدت تابش خورشیدی در ایستگاه یزد پرداخته است. لازم به ذکر است که مطالعات زیادی در خصوص برآورد شدت تابش خورشیدی در برخی ایستگاه‌های ایران انجام شده است که تنها از داده‌های هواشناسی استفاده شده است. تنها مطالعه موجود که تابش خورشیدی در کرمان را با استفاده از مشخصات هندسی، نجومی، جغرافیایی و هواشناسی برآورد کرده است مربوط به مطالعه (Saffaripour and Mehrabian, 2009) می‌باشد. به منظور بررسی دقیق‌تر تأثیر نسبی این متغیرها بر روی شدت تابش خورشیدی، ترکیبی از داده‌های هواشناسی و داده‌های هندسی، جغرافیایی و نجومی از روش آنالیز مؤلفه‌های اصلی استفاده گردید و با انتخاب چند مؤلفه اصلی اول، سایر متغیرها از محاسبات بعدی حذف شدند و با کاربرد رگرسیون خطی چند متغیره و مؤلفه‌های اصلی ایجاد شده یک مدل رگرسیونی برای محاسبه شدت تابش خورشیدی بدست آمد.

مواد و روش‌ها

معرفی منطقه مورد مطالعه و داده‌های مورد استفاده

در مطالعه حاضر ایستگاه سینوپتیک یزد که در مرکز ایران با عرض و طول جغرافیایی به ترتیب ۳۱/۸۹۷۴ درجه شمالی و ۵۴/۳۵۶۹ درجه شرقی در ارتفاع ۱۲۱۶ متری از سطح دریا قرار گرفته است، انتخاب گردید. داده‌های مورد نیاز در این تحقیق شامل دمای کمینه، دمای بیشینه، میانگین دما، رطوبت نسبی، ساعات آفتابی و شدت تابش خورشیدی

همبستگی‌ها موجود در بین داده‌ها برای تحلیل مؤلفه‌های اصلی مناسب خواهند بود. این عامل به کمک ضرایب همبستگی ساده و جزئی طبق معادله ۶ محاسبه می‌شود.

$$KMO = \frac{\sum_{i=1}^p \sum_{j=1}^p r_{ij}^2}{\sum_{i=1}^p \sum_{j=1}^p r_{ij}^2 + \sum_{i=1}^p \sum_{j=1}^p a_{ij}^2} \quad (6)$$

که در آن ضریب همبستگی ساده بین متغیرهای i و j و a_{ij} ضریب همبستگی جزئی بین متغیرهای i و j برابر با تعداد متغیرها می‌باشد. با انتخاب چند مؤلفه اصلی اول، سایر مؤلفه‌ها از محاسبات بعدی حذف می‌شوند. از نمودار واریزه‌ای (اسکری پلات) که در آن مقادیر ویژه در مقابل شماره مؤلفه‌ها رسم می‌شود، برای تشخیص آستانه حذف استفاده می‌شود. در این روش مرز بین مؤلفه‌های اصلی و غیر اصلی محلی است که نمودار میل به خطی شدن به صورت افقی می‌نماید. به منظور دستیابی به ماتریسی ساده و از نظر تئوریک معنادار و قابل تفسیر از چرخش عامل‌ها استفاده می‌شود. عامل‌ها به دور مبدا چرخش داده می‌شوند تا موقعیت جدیدی کسب کنند. روش چرخش واریماکس به دنبال مختصر کردن و ساده‌سازی ستون‌های ماتریس عاملی است. در نتیجه روش واریماکس استفاده وسیع‌تری در بین محققین دارد و یکی از انواع چرخش‌های عمودی است (Johnson and Wichern, 1982).

مدل رگرسیون خطی چند متغیره

مدل رگرسیون خطی چندگانه در فرم ماتریسی را می‌توان به صورت معادله Y نشان داد.

$$Y = X\beta + \varepsilon \quad (7)$$

که در آن β ماتریس ضرایب رگرسیون، ε ماتریس خطای برازش، Y ماتریس پاسخ (متغیر وابسته) و X ماتریس متغیرهای مستقل می‌باشد. با حل رابطه بر حسب β خواهیم داشت:

$$\beta = (X'X)^{-1}(X'Y) \quad (8)$$

که در آن ترانهاده ماتریس X است. برای محاسبه وارون $(X'X)$ لازم است قبل از ساخت مدل رگرسیونی، همبستگی

جاییکه R_a تابش فزاینده بر حسب مگاژول بر مترمربع بر روز، I_{gs} ثابت خورشیدی و برابر با 1367 وات بر متر مربع، J شماره روز از سال، d_r فاصله نسبی زمین تا خورشید، φ عرض جغرافیایی بر حسب درجه، δ زاویه میل خورشیدی بر حسب درجه، w_s زاویه ساعتی خورشیدی بر حسب درجه و N حداکثر ساعات روشنایی روز است.

جدول ۱- شاخص‌های آماری متغیرهای مورد بررسی

Table 1- Statistical indices of the investigated variables

Parameters	Min	Max	Mean	SD
R_s ($Mjm^{-2}day^{-1}$)	4.0	30.6	19.3	6.2
R_a ($Mjm^{-2}day^{-1}$)	31.6	33.0	32.3	0.4
n/N	0.01	1	0.7	0.2
RH (%)	5	95.4	27.3	17.7
Tmax-Tmin ($^{\circ}C$)	2.8	28	13.9	3.1
T_{mean} ($^{\circ}C$)	-3.6	45.6	28	9.3
δ ($^{\circ}$)	0.00031	0.035	0.0126	0.013
Dr ($^{\circ}$)	1.032	1.033	1.032	0.00006

روش آنالیز مؤلفه‌های اصلی

در روش آنالیز مؤلفه‌های اصلی P متغیر اصلی همبسته به p مؤلفه غیرهمبسته یا متعامد تبدیل می‌شوند. با اعمال PCA متغیرهای ورودی اصلی به متغیرهای جدید که بدون همبستگی می‌باشند، تبدیل می‌شوند. مؤلفه‌های ایجاد شده ترکیبی خطی متغیرهای اصلی می‌باشند (Liu et al., 2003). اگر در P متغیر اصلی تنها مقداری همبستگی وجود داشته باشد انجام روش PCA می‌تواند مفید باشد. به همین دلیل اولین گام برای اجرای این روش تعیین و تشخیص مناسب بودن داده‌ها است. روش‌هایی وجود دارد که محقق از طریق آنها می‌تواند این مطلب را بیازماید. از جمله این روش‌ها می‌توان به محاسبه فاکتور KMO ^۱ اشاره کرد. مقدار این عامل همواره بین صفر و یک در نوسان است. در صورتیکه KMO کمتر از $0/5$ باشد، داده‌ها برای روش تحلیل مؤلفه‌های اصلی مناسب نخواهند بود. اگر مقدار آن بین $0/5$ تا $0/69$ باشد، داده‌ها متوسط و اگر مقدار این شاخص، بزرگتر از $0/7$ باشد

¹ Kaiser-Meyer-Olkin Measure of sampling adequacy

متغیرهای توصیفی از اهمیت خاصی برخوردار است. در حقیقت همخطی در مدل رگرسیونی به وجود همبستگی بین متغیرهای مستقل اشاره دارد و نتایج مدل رگرسیون چندگانه را غیرمعتبر می‌کند. به منظور شناسایی همخطی از معیار عامل تورم واریانس^۲ برای مدل رگرسیونی استفاده می‌شود. مقدار VIF همواره ۱ یا بیشتر از ۱ است. شرایطی که تمام مقادیر VIF در مدل، نزدیک ۱ باشند، همخطی وجود ندارد و مقدار بیشتر از ۱۰ برای این شاخص نشان از همخطی جدی بوده و باعث عدم اعتماد به نتایج می‌شود. برای رفع چنین حالتی می‌توان از راهکارهای مختلفی کمک گرفت. استفاده از تکنیک تحلیل مؤلفه‌های اصلی به منظور کاهش بعد مسئله و استفاده از مؤلفه‌های اصلی در ایجاد مدل رگرسیونی یکی از این روشهاست. از آنجایی که این مؤلفه‌ها دارای کمترین همبستگی هستند، مشکل همخطی در مدل توسط این مؤلفه‌ها رفع خواهد شد (Montgomery et al., 2012). در این تحقیق نیز بعد از ساخت مدل‌های رگرسیون خطی برای بررسی رابطه همخطی بین متغیرهای ورودی مدل، VIF محاسبه گردید.

نتایج و بحث

پیش‌پردازش متغیرهای ورودی به مدل رگرسیونی با استفاده از PCA

برای بررسی امکان تجزیه و تحلیل مؤلفه‌های اصلی، از آزمون بارتلت^۳ استفاده شد. مقدار KMO برابر با ۰/۷۲ امکان اجرای آنالیز مؤلفه‌های اصلی را تأیید کرد. برای اجرای این روش پس از استاندارد کردن متغیرهای ورودی، ماتریس متقارن همبستگی تشکیل شد. نتایج حاصل از ضرایب همبستگی بین مقادیر شدت تابش خورشیدی و برخی متغیرهای تأثیرگذار (جدول ۲) نشان داد که شدت تابش خورشیدی با همه متغیرهای مورد بررسی همبستگی مثبت معنی‌دار دارد بجز متغیرهای رطوبت نسبی و زاویه میل خورشیدی که رابطه معنی‌دار منفی مشاهده گردید. جدول ۳ مقادیر ویژه را نشان می‌دهد. مطابق این جدول، دو مؤلفه اول در مجموع حدود ۸۵ درصد از پراکندگی داده‌های اصلی را

بین متغیرهای مستقل را از بین برد. در این خصوص روش مناسب استفاده از تحلیل مؤلفه‌های اصلی روی متغیرهای مستقل ورودی به مدل است. ورود متغیرها به مدل رگرسیونی به صورت مرحله‌ای، از مهمترین متغیر تا کم‌اهمیت‌ترین آنها، صورت می‌گیرد. معیار میزان اهمیت متغیر در مدل، مقدار سطح معنی‌دار یا آماره t متناظر با آن در جدول آزمون معنی‌داری متغیرها می‌باشد. جهت مدل‌سازی، مقادیر بدست آمده از PC1 و PC2 و غیره به عنوان متغیرهای مستقل (برآورد کننده) و مقادیر شدت تابش خورشیدی به عنوان متغیر وابسته وارد مدل رگرسیون خطی چندگانه گردیدند. در این پژوهش برای انجام تجزیه تحلیل مؤلفه‌های اصلی و مدل‌سازی شدت تابش خورشیدی از نرم‌افزار SPSS ۱۶/۰ استفاده شد. در این مطالعه از روش رگرسیون مؤلفه اصلی^۱ استفاده گردید. این روش یک تکنیک تحلیل رگرسیونی است که بر اساس تجزیه و تحلیل مؤلفه اصلی (PCA) است. به طور خاص، PCR برای تخمین ضرایب مجهول رگرسیون در یک مدل رگرسیون خطی استاندارد استفاده می‌شود.

شاخص‌های ارزیابی مدل‌ها

در مطالعه حاضر به منظور ارزیابی معادلات تجربی برآورد شدت تابش خورشیدی روزانه از دو شاخص آماری شامل ریشه میانگین مربع خطا (RMSE)، و ضریب تبیین (R^2) استفاده شدند. مقادیر کمتر شاخص اول، پیش‌بینی بهتر مدل‌ها را نشان می‌دهد. ضریب تبیین بین ۰ و ۱ بوده و مقادیر بیشتر از این شاخص کارایی بهتر مدل را نمایان می‌سازد.

$$RMSE = \sqrt{\sum_{i=1}^N (RS_{ic} - RS_{im})^2} \quad (9)$$

$$R^2 = \left(\frac{\sum_{i=1}^N (RS_{ic} - RS_{cave})(RS_{im} - RS_{mave})}{\sqrt{\left(\sum_{i=1}^N (RS_{ic} - RS_{cave})^2 \right) \times \left(\sum_{i=1}^N (RS_{im} - RS_{mave})^2 \right)}} \right)^2 \quad (10)$$

جایی که N تعداد مشاهدات، RS_{ic} و RS_{im} به ترتیب مقادیر مشاهده شده شدت تابش خورشیدی و مقادیر برآورد شده شدت تابش خورشیدی و RS_{cave} و RS_{mave} به ترتیب میانگین مقادیر برآورد شده و مشاهده شده شدت تابش خورشیدی می‌باشد. در رگرسیون خطی، شرط مستقل بودن

³ Bartlett's Test of Sphericity

¹ Principal Component Regression (PCR)

² Variance Inflation Factor (VIF)

$$PCI = 0.910 \times T_{\min} + 0.963 \times T_{\max} + 0.949 \times T_{\text{mean}} - 0.801 \times RH + 0.6 \times \frac{n}{N} - 0.147 \times d_r + 0.349 \times \delta + 0.343 \times R_a \quad (11)$$

$$PC2 = -0.201 \times T_{\min} - 0.190 \times T_{\max} - 0.198 \times T_{\text{mean}} + 0.237 \times RH - 0.804 \times \frac{n}{N} - 0.986 \times d_r + 0.936 \times \delta + 0.939 \times R_a \quad (12)$$

چون در تشکیل هر مؤلفه از تمام متغیرهای اولیه استفاده شده است، از روش چرخش واریماکس برای تفسیر ساده‌تر استفاده شد و همبستگی دو مؤلفه اول نسبت به متغیرهای مورد استفاده و مقدار شدت تابش خورشیدی بررسی گردید. این چرخش سبب می‌شود وابستگی عوامل بارگذاری شده با مؤلفه‌های اصلی آشکارتر شده و تشخیص اینکه تغییرات هر یک از مؤلفه‌های اصلی متأثر از کدام عوامل بارگذاری است، آسان‌تر باشد. با توجه به نتایج جدول ۵، ضریب همبستگی مؤلفه‌های اول و دوم با مقدار شدت تابش خورشیدی به ترتیب ۰/۸۹۳ و ۰/۱۶۸- است. بنابراین مؤلفه اول نسبت به مؤلفه دوم با مقدار شدت تابش خورشیدی همبستگی بیشتری دارد و در نتیجه متغیرهایی که در این مؤلفه از نظر مقدار عددی سهم بیشتر (بارگذاری بزرگتری) دارند، همبستگی بیشتری با شدت تابش خورشیدی خواهند داشت. یا به عبارت دیگر تأثیر متغیرهای هواشناسی مانند رطوبت نسبی، دمای کمینه و بیشینه بر شدت تابش خورشیدی بسیار بیشتر است. مقدار دوران شده بارگذاری، در مؤلفه اول نشان می‌دهد که نسبت ساعات آفتابی، تابش فرازمینی، زاویه میل خورشیدی و فاصله نسبی زمین تا خورشید با شدت تابش خورشیدی همبستگی کمتری دارند به گونه‌ای که همبستگی تابش فرازمینی و زاویه میل خورشیدی در جهت مثبت و همبستگی فاصله نسبی زمین تا خورشید و نسبت ساعات آفتابی در جهت منفی است و همبستگی بقیه عوامل (رطوبت نسبی، دمای کمینه و بیشینه و میانگین) در این مؤلفه بسیار زیاد است. در مقابل، در مؤلفه دوم بیشترین بارگذاری مربوط به تابش فرازمینی، زاویه میل خورشیدی، نسبت ساعات آفتابی و فاصله نسبی زمین تا خورشید می‌باشد. با توجه به نتایج این قسمت، آنالیز مؤلفه اصلی داده‌های ورودی را به دو مؤلفه بر اساس ورودی داده‌های هواشناسی و ورودی داده‌های هندسی، جغرافیایی و

بین می‌کند. بنابراین، در مطالعه حاضر می‌توان دو مؤلفه اول را به عنوان مؤلفه اصلی انتخاب کرد. چراکه در نمودار واریزه‌ای نیز تغییرات مقادیر ویژه با شماره مؤلفه‌ها برای مؤلفه‌های بزرگتر از ۲ خطی افقی را تشکیل می‌دهد. در جدول ۴ مقادیر بردارهای ویژه مربوط به متغیرهای مورد بررسی ذکر شده است که ضرایب هر مؤلفه را برای محاسبه آن‌ها تعیین می‌کند.

جدول ۲- همبستگی متغیرهای تأثیرگذار بر مقدار تابش

خورشیدی

Table 2- Correlation of parameters affecting the amount of solar radiation

	Ra	n/N	RH	$\frac{T_{\max}-T_{\min}}$	T_{mean}	δ	d_r
Rs	0.07**	0.7**	-0.6**	0.3**	0.8**	-0.06*	0.3**

جدول ۳- مشخصات مقادیر ویژه حاصل از متغیرها و درصد

واریانس

Table 3- Characteristics of eigenvalues from parameters and percentage of variance

Component	Eigenvalues		
	Sum	Percentage of variance	Cumulative percentage
1	3.920	49.03	49.03
2	2.909	36.367	85.370
3	0.793	9.913	95.283
4	0.343	4.292	99.575
5	0.028	0.351	99.926
6	0.006	0.074	100
7	3.29×10^{-10}	4.11×10^{-9}	100
8	7.42×10^{-18}	9.27×10^{-17}	100

جدول ۴- مقادیر بردار ویژه مربوط به متغیرهای مورد بررسی

Table 4- Eigenvector values related to the examined parameters

Parameters	Component 1	Component 2
T_{\min}	0.910	-0.201
T_{\max}	0.963	-0.190
T_{mean}	0.949	-0.198
RH	-0.801	0.237
n/N	0.600	-0.804
d_r	-0.147	-0.986
δ	0.349	0.936
Ra	0.343	0.939

برای تشکیل هر مؤلفه باید مقادیر متغیرها را در بردارهای ویژه مربوط به هر متغیر ضرب و حاصل را با هم جمع کرد. در نتیجه، مؤلفه‌هایی حاصل می‌شود که می‌توان از آنها به جای متغیرهای اولیه به عنوان ورودی به مدل رگرسیون استفاده کرد (معادله‌های ۱۱ و ۱۲).

جدول ۶- نتایج نهایی ضرایب رگرسیونی رابطه شدت تابش خورشیدی و مؤلفه‌های اصلی

Table 6- The final results of the regression coefficients of the relationship between solar radiation intensity and the main components

Model	coefficients			t	Sig	VIF
	Non-standard	Standard				
	SD	B	Beta			
constant	0.06	19.44	-	294.9	0.00	
PC1	0.06	5.47	0.83	83.04	0.00	1.04
PC2	0.06	-1.03	-0.168	15.65	0.00	1.01

ارزیابی کارائی مدل رگرسیون مؤلفه اصلی

به منظور مقایسه نتایج روش رگرسیون مؤلفه اصلی با مقادیر حاصل از روش رگرسیون خطی تمامی متغیرهای مورد استفاده در این پژوهش وارد مدل رگرسیونی گردید و معادله رگرسیون خطی چند متغیره حاصل گردید. با توجه به این معادله، آماره t برای متغیرهای تابش فرازمینی و دمای میانگین در سطح ۵ درصد معنی‌دار نبود و در نتیجه از معادله رگرسیون خارج شدند. با توجه به اینکه آنالیز مؤلفه‌های اصلی دو عامل (عامل هواشناسی و عامل هندسی) را شناسائی نمود، ترکیب‌های مختلفی از ورودی‌ها به مدل رگرسیونی وارد شد تا در نتیجه ۵ مدل رگرسیون خطی چند متغیره بر اساس ورودی‌های مختلف بدست آمد (معادله‌های ۱۴ تا ۱۸). نتایج عامل تورم واریانس این مدل‌ها نشان دهنده نامناسب بودن برخی مدل‌های رگرسیون خطی (مدل‌های ۲، ۳ و ۴) بدلیل وجود همبستگی بین متغیرهای مدل و کسب VIF بیشتر از ۱۰ می‌باشد که نتایج مدل رگرسیون را غیرمعتبر می‌کند. تنها دو مدل شماره ۱ و ۵ با کسب VIF کمتر از ۱۰ مدل‌های مناسب تشخیص داده شده و برای مقایسه با مدل رگرسیون مؤلفه اصلی مورد آزمون قرار گرفتند.

$$R_s = 12.716 + 0.379 \times T_{\min} - 0.092 \times RH + 0.19 \times (T_{\max} - T_{\min})$$

$$R = 0.786$$

$$VIF(T_{\max} - T_{\min}) = 1.01$$

$$VIF(T_{\min}) = 1.52$$

$$VIF(RH) = 1.52$$
(۱۵)

$$R_s = 12.716 + 0.115 \times T_{\min} + 0.278 \times T_{\max} - 0.08 \times RH$$

$$R = 0.786$$

$$VIF(T_{\min}) = 11.97$$

$$VIF(T_{\max}) = 17.14$$

$$VIF(RH) = 2.85$$
(۱۶)

نجومی وابسته به فاصله و زاویه به منظور تخمین شدت تابش خورشیدی تقسیم می‌کند.

ساخت مدل رگرسیونی

برای مدل‌سازی، مقادیر مؤلفه‌های بدست آمده PC1 و PC2 به عنوان متغیرهای مستقل و مقادیر تابش خورشیدی (R_s) به عنوان متغیر وابسته، به مدل رگرسیونی خطی چندگانه وارد شد. بررسی رابطه بین مقادیر تابش خورشیدی (R_s) و مقادیر مؤلفه‌های اصلی بدست آمده از PC1 و PC2 نشان داد که اثر متغیرهای مورد نظر بر تابش خورشیدی معنی‌دار ($p < 0.001$) و ضریب تبیین آن ۰/۸۹ درصد است. جدول ۶، نتایج ضرایب رگرسیونی رابطه تابش خورشیدی و مقادیر مؤلفه‌های اصلی به دست آمده از PC1 و PC2 را نشان می‌دهد. بر اساس نتایج این جدول و شاخص عامل تورم واریانس کمتر از ده برای هر دو مؤلفه، رابطه شدت تابش خورشیدی و مقادیر مؤلفه‌های اصلی را می‌توان به صورت معادله ۱۳ نوشت.

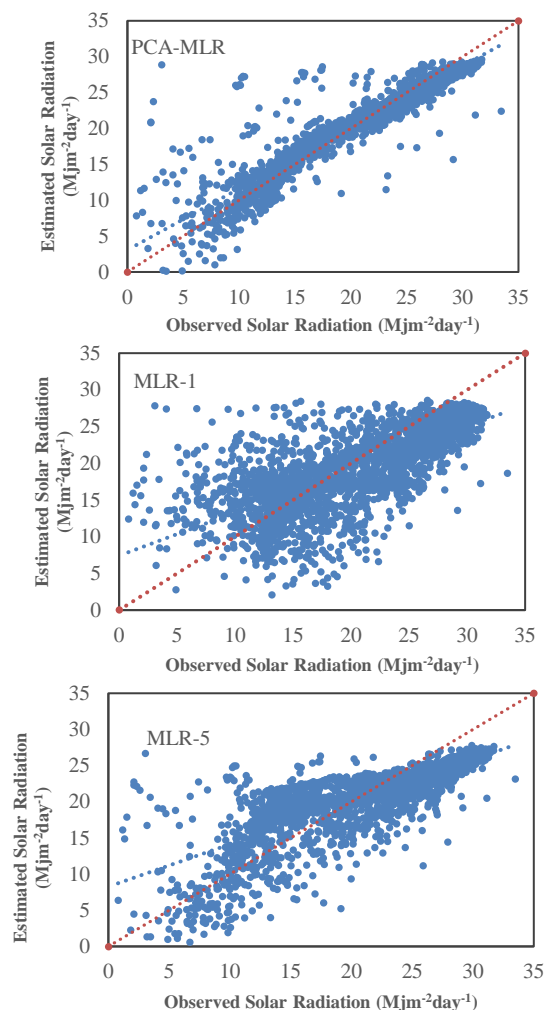
$$R_s = 19.44 + 5.47PC1 - 1.03PC2 \quad (13)$$

$$R_s = 19.44 + 5.47 \times 0.910 \times T_{\min} + 0.963 \times T_{\max} + 0.949 \times T_{\text{mean}} - 0.801 \times RH + 0.600 \times \frac{n}{N} - 0.147 \times d_r + 0.349 \times \delta + 0.343 \times R_a + 1.03 \times 0.201 \times T_{\min} - 0.190 \times T_{\max} - 0.198 \times T_{\text{mean}} + 0.237 \times RH - 0.804 \times \frac{n}{N} - 0.986 \times d_r + 0.936 \times \delta + 0.939 \times R_a$$
(۱۴)

که در آن R_s شدت تابش خورشیدی بر حسب مگاژول بر مترمربع بر روز است.

جدول ۵- بردارهای ویژه بعد از دوران با روش واریماکس
Table 5- Eigenvectors after rotation with varimax method

Parameters	Components after rotation	
	Component 1	Component 2
T_{\min}	0.913	0.053
T_{\max}	0.965	0.077
T_{mean}	0.952	0.066
RH	-0.825	0.013
n/N	0.654	0.058
d_r	0.120	-0.990
δ	0.09	0.995
R_a	0.084	0.995
R_s	0.893	-0.168



شکل ۱- مقادیر مشاهده شده و برآورد شده تابش خورشیدی در مرحله آزمون

Figure 1- Observed and estimated values of solar radiation in the test stage

به نظر می‌رسد در ایستگاه یزد مدل‌های رگرسیونی مبتنی بر دما نمی‌توانند بخوبی شدت تابش خورشیدی را برآورد کنند و در مقابل داده‌های ساعات آفتابی و تابش فرازمینی برآورد بهتری از تابش خورشیدی خواهند داشت. (Besharat et al., 2013) نیز در برآورد تابش خورشیدی ماهانه، روش‌های مبتنی بر ساعات آفتابی را در ایستگاه یزد مناسب دانستند. همچنین Fooladmand and Karimi, (2017) نیز نشان دادند که مدل‌های وابسته به دمای هوا برای تخمین تابش خورشیدی در منطقه شیراز مناسب نیستند و استفاده از

$$R_s = -278178.662 + 269303.969 \times dr + 1139.791 \times \delta + 14.619 \times \frac{n}{N}$$

$$R=0.946$$

$$VIF(\delta)=22.59$$

$$VIF(dr)=22.4$$

$$VIF\left(\frac{n}{N}\right)=1.18$$

$$R_s = -369895.438 + 358102.244 \times dr + 11516.102 \times \delta$$

$$R=0.841$$

$$VIF(\delta)=19.04$$

$$VIF(dr)=19.04$$

$$R_s = 76.093 - 2.293 \times R_a + 22.919 \times \frac{n}{N}$$

$$R=0.757$$

$$VIF(R_a)=1.06$$

$$VIF\left(\frac{n}{N}\right)=2.38$$

جدول ۷ معیارهای ارزیابی محاسبه شده جهت مقایسه مدل‌ها در ایستگاه یزد در دو مرحله آموزش و آزمون بیان شده است.

جدول ۷- معیارهای ارزیابی محاسبه شده جهت مقایسه مدل‌ها در ایستگاه یزد در دو مرحله آموزش و آزمون

Table 7- Evaluation criteria calculated to compare models at Yazd station in two stages of training and testing

		PCR	MLR-1	MLR-5
Train	RMSE	0.061	0.066	0.052
	R	0.901	0.780	0.902
Test	RMSE	0.028	0.076	0.070
	R	0.911	0.789	0.852

شکل ۱ نیز مقادیر مشاهده شده شدت تابش خورشیدی و مقادیر برآورد شده از مدل‌های پیشنهادی در مرحله آزمون را نشان می‌دهد. همانطور که نتایج ارزیابی مدل‌ها در مرحله آزمون نشان می‌دهد، معادله استخراج شده توسط آنالیز مؤلفه‌های اصلی بهترین نتیجه (ضریب تبیین = ۰/۹۱۱ و ریشه میانگین مربع خطا = ۰/۰۲۸) را به دنبال داشت. به نظر می‌رسد مدل حاصل از آنالیز مؤلفه اصلی با وزن‌دهی مناسب به متغیرهای تأثیرگذار با دقت بیشتری قادر به پیش‌بینی شدت تابش خورشیدی می‌باشد. کمترین ضریب تبیین و بیشترین ریشه میانگین مربع خطا در مدل رگرسیون خطی چند متغیره شماره ۱ که بر اساس تفاوت دمای حداقل و حداکثر بدست آمد، مشاهده گردید.

در ایستگاه یزد شامل دمای کمینه، دمای بیشینه، دمای میانگین، رطوبت نسبی و ساعات آفتابی نسبت به متغیرهای تابش فرازمینی، فاصله نسبی زمین تا خورشید و زاویه میل خورشیدی تأثیر بیشتری بر شدت تابش خورشیدی دارند. همچنین نتایج عامل تورم واریانس در مدل‌های رگرسیونی نشان داد که برخی مدل‌های رگرسیونی با وجود ضریب تبیین مناسب بدلیل مشکل هم‌خطی در متغیرهای ورودی مدل، نمی‌توانند قابل اطمینان باشند. تکنیک مؤلفه‌های اصلی با کاهش بعد مسئله و استفاده از مؤلفه‌های اصلی در ایجاد مدل رگرسیونی مشکل هم‌خطی این مدل‌ها را مرتفع ساخت. نتایج ارزیابی مدل‌ها نیز حاکی از توانایی روش رگرسیون مؤلفه اصلی در پیش‌بینی تابش خورشیدی نسبت به روش رگرسیون خطی چند متغیره است. می‌توان نتیجه گرفت که روش تحلیل مؤلفه اصلی باعث بهبود نتایج مدل‌سازی می‌شود. این نتایج با نتایج سایر پژوهشگران همخوانی دارد، از جمله پژوهش (Mohammadi et al., (2018), Noori et al., (2010) و (Mohammadi and Emamgholizadeh (2017) که همگی ادعان داشتند روش تحلیل مؤلفه اصلی باعث بهبود نتایج مدل‌سازی می‌شود. پیشنهاد می‌گردد این روش برای ایستگاه‌های هواشناسی با اقلیم‌های متفاوت از ایستگاه یزد نیز انجام شده و نتایج با یافته‌های این تحقیق مورد مقایسه قرار گیرد.

منابع

- Abdelhafidi, N., Bachari, N.E.I., Abdelhafidi, Z., 2021. Estimation of solar radiation using stepwise multiple linear regression with principal component analysis in Algeria. *Meteorology and Atmospheric Physics*, 133(2): 205-216.
- Angstrom, A. 1924. Solar and terrestrial radiation. *Quarterly Journal of the Royal Meteorological Society*, 50: 121-125.
- Almorox, J., Benito, M., Hontoria, C., 2005. Estimation of monthly Angstrom-PreScott equation coefficients from measured daily data in Toledo, Spain. *Renewable Energy*, 30, 931-936.
- Almorox, J., Bocco, M., Willington, E., 2013. Estimation of daily global solar radiation from measured temperatures at Cañada de Luque, Córdoba, Argentina. *Renewable Energy*, 60, 382-387.

ساعات آفتابی برای تخمین تابش خورشیدی در این منطقه ضرورت دارد.

نتیجه‌گیری

با توجه به اهمیت تابش خورشیدی در برنامه‌ریزی و مدیریت منابع آب و لزوم استفاده از روش‌های نوین در برآورد شدت تابش خورشیدی، در این پژوهش از روش‌های تحلیل مؤلفه اصلی و رگرسیون خطی چند متغیره برای پیش‌بینی شدت تابش خورشیدی استفاده شد. شدت تابش خورشیدی با همه متغیرهای مورد بررسی همبستگی مثبت معنی‌دار نشان داد؛ بجز متغیرهای رطوبت نسبی و زاویه میل خورشیدی که رابطه معنی‌دار منفی مشاهده گردید. با افزایش بخار آب و رطوبت در هوا شدت تابش خورشیدی کمتری به سطح زمین می‌رسد. همچنین با افزایش فاصله نسبی زمین تا خورشید نیز انرژی دریافتی و شدت تابش خورشیدی کمتر خواهد شد. (Seyedian et al., (2017) نیز نشان دادند در ایستگاه‌های اصفهان و زاهدان که اقلیمی مشابه با اقلیم ایستگاه یزد دارند، متغیر رطوبت نسبی متغیری تأثیرگذار بر شدت تابش خورشیدی می‌باشد و دارای همبستگی منفی می‌باشد. هر چند در ایستگاه‌های ارومیه و رامسر رطوبت نسبی جزو متغیرهای کم اهمیت بر شدت تابش خورشیدی تعیین گردید. (Zeng and Qiao, (2013) نیز نشان دادند که در تمام ایستگاه‌ها رطوبت نسبی با شدت تابش خورشیدی رابطه عکس دارد. هم‌چنین نتایج نشان داد که روش آنالیز مؤلفه‌های اصلی توانست بخوبی اهمیت نسبی متغیرهای مؤثر بر شدت تابش خورشیدی را بیان کند. آنالیز مؤلفه اصلی داده‌های ورودی را به دو مؤلفه بر اساس ورودی داده‌های هواشناسی و ورودی داده‌های هندسی وابسته به فاصله و زاویه به منظور تخمین شدت تابش خورشیدی تقسیم می‌کند. نتایج با یافته‌های بسیاری از محققین همخوانی دارد چرا که بسیاری از معادلات رگرسیونی پیشنهاد شده برای برآورد شدت تابش خورشیدی نیز بر اساس داده‌های دما، داده‌های تابش خورشیدی یا داده‌های ترکیبی هواشناسی گروه‌بندی می‌شوند (Fooladmand and Karimi, 2017; Seyedian et al., 2017; Piri and Kisi, 2015; Mecibah et al., 2014; Citakoglu, 2015). بر اساس این روش، متغیرهای هواشناسی

- Fooladmand, H.R., Karimi, F. 2017. Calibration and Evaluation of Different Estimation Models of Daily Solar Radiation in Seasonally and Annual Time Steps in Shiraz Region. *Journal of Water and Soil*, 9(34): 137-150. (In Farsi)
- Hunt, L.A., Kuchar, L., Swanton, C.J., 1998. Estimation of solar radiation for use in crop modeling. *Journal of Agricultural Meteorology*, 91, 293-300.
- Iziomon, M.G., Mayer, H., 2002. Assessment of some global solar radiation parameterizations. *Journal of Atmospheric and Solar-Terrestrial Physics*, 64, 1631-1643.
- Janjai, S., Sricharoen, K., Pattarapanitchai, S., 2011. Semi-empirical models for the estimation of clear sky solar global and direct normal irradiances in the tropics. *Applied Energy*, 88, 4749-4755.
- Jiang, Y., 2009. Estimation of monthly mean daily diffuse radiation in China. *Applied Energy*, 86, 1458-1464.
- Johnson R.A., Wichern, D.W. 1982. *Applied multivariate statistical analysis*. Prentice-Hall Inc., Englewood Cliffs, SA, 590 p.
- Li, H., Ma, W., Lian, Y., Wang, X., Zhao, L. 2011. Global solar radiation estimation with sunshine duration in Tibet, China. *Renewable Energy*, 36(11):3141-3145.
- Liu, C.W., Lin, K.H., Kuo, Y.M. 2003. Application of factor analysis in the assessment of groundwater quality in a blackfoot disease area in Taiwan. *Science of the Total Environment* 313: 77-89.
- Liu, J., Linderholm, H., Chen, D., Zhou, X., Flerchinger, G.N., Yu, Q., Du, J., Wu, D., Shen, Y., Yang, Z., 2015. Changes in the relationship between solar radiation and sunshine duration in large cities of China. *Energy* 82, 589-600
- Mecibah, M. S., Boukelia, T. E., Tahtah, R., Gairaa, K. 2014. Introducing the best model for estimation the monthly mean daily global solar radiation on a horizontal surface (case study: Algeria). *Renewable and Sustainable Energy Reviews*, 36:194-202.
- Medeiros, F.Jd., Silva, CM., Bezerra, B.G. 2017. Calibration of ångströmprescott equation to estimate daily solar radiation on rio Grande do Norte state, Brazil. *Revista Brasileira de Meteorologia*, 32(3):409-416.
- Menges, H.O., Ertekin, C., Sonmete, M.H., 2006. Evaluation of global solar radiation models for Konya, Turkey. *Energy Conversion and Management*, 47, 3149-3173.
- Mohan, S., Arumugam, N. 1996. Relative importance of meteorological variables in evapotranspiration: Factor analysis approach. *Journal of Water Resources Management*, 10: 1-20.
- Avazpour, S., Bakhtiari, B., Qaderi, K., 2019. Performance evaluation of Neural Network and Multivariate Regression Methods for Estimation of Total Solar Radiation at several stations in Arid and Semi-arid Climates. *Journal of Soil and Water Research*, 50(8): 1855-1869. (In Farsi).
- Besharat, F., Dehghan, A. and Faghieh, A.R., 2013. Empirical models for estimating global solar radiation: a review and case study, *Renewable and Sustainable Energy Reviews*, 21: 798-821.
- Ebrahimi, H., Agha-Shariatmadari, Z., Hijabi, S., Nazi Qomshloo, A. 2020. Evaluation of the efficiency of Neuro-Fuzzy Adaptive Inference System (ANFIS) in estimating total solar radiation. *Agricultural Meteorology*, 8 (1): 3-14. (In Farsi)
- Bakirci, K., 2009. Models of solar radiation with hours of bright sunshine: a review. *Renewable and Sustainable Energy Reviews*, 13, 2580-2588.
- Bakirci, K., 2015. Models for the estimation of diffuse solar radiation for typical cities in Turkey. *Energy* 82:827-838.
- Besharat, F., Dehghan, A.A., Faghieh, A.R., 2013. Empirical models for estimating global solar radiation: a review and case study. *Renewable and Sustainable Energy Reviews*, 21, 798-821.
- Citakoglu, H., 2015. Comparison of artificial intelligence techniques via empirical equations for prediction of solar radiation. *Computers and Electronics in Agriculture* 118:28-37.
- Chelbi, M., Gagnon, Y., Waewsak, J. 2015. Solar radiation mapping using sunshine duration-based models and interpolation techniques: application to Tunisia. *Energy Conversion and Management*, 101:203-215.
- Chukwujindu, N.S., 2017. A comprehensive review of empirical models for estimating global solar radiation in Africa, *Renewable and sustainable energy reviews*, 78, 955-995.
- De Souza, J. L., Lyra, G. B., Dos Santos, C.M., Junior, R.A.F., Tiba, C., Lyra, G. B., Lemes, M. A. M. 2016. Empirical models of daily and monthly global solar irradiation using sunshine duration for Alagoas state, northeastern Brazil. *Sustainable Energy Technologies and Assessments*, 14:35-45.
- Deverel, S.J. 1989. Geostatistical and principal-component analyses of groundwater chemistry and soil-salinity data, San Joaquin Valley, California. *Regional Characterization of Water Quality*, 182:11-18.
- Duffie, J.A., Beckman, W.A., 1991. *Solar Engineering of Thermal Processes*. Wiley, New York.
- Ertekin, C., Evrendilek, F., 2007. Spatio-temporal modeling of global solar radiation dynamics as a function of sunshine duration for Turkey. *Journal of Agricultural Meteorology*, 145, 36-47.

- estimating total solar radiation. *Natural Geography Research*, 51 (2): 353-372. (In Farsi)
- Saffaripour, M., Mehrabian, M. 2009. Predicting the total amount of solar radiation in Kerman using geometric, astronomical, geographical and meteorological characteristics. *Sharif*, 51 (1): 3-13. (In Farsi)
- Seifi, A., Mirlatifi, S. M., Riyahi, H. 2011. Evaluation and monitoring of meteorological stations by principal component analysis and factor analysis. (Case study: Kerman station). *Irrigation and drainage of Iran*, 1(5):30-42. (In Farsi)
- Seyedian, S.M., Farasati, M., Rouhani, H., Heshmatpour, A. 2017. Solar Radiation Prediction Using Metrological Parameters, *Iran-Water Resources Research*, 13(1):88-100. (In Farsi)
- Sheikholeslami, N., Ghahraman, B., Mosaedi, A., Davary, K., Mohejerpour, M. 2014. Estimating Reference Evapotranspiration by Using Principal Component Analysis (PCA) and The Development of a Regression Model (MLR-PCA) (Case Study: Mashhad Station). *Journal of Water and Soil*, 28(2): 420-429. (In Farsi)
- Skeiker, K. 2006. Correlation of global solar radiation with common geographical and meteorological parameters for Damascus province, Syria, *Energy Conversion and Management, Mgmt*, 47: 331-345.
- Soltani-Gerdefaramarzi, S., Ghasemi, M. and Gheysouri, M., 2021. Pollution, human health risk assessment and spatial distribution of toxic metals in urban soil of Yazd City, Iran. *Environmental Geochemistry and Health*, 43(9): 3469-3484.
- Sonmete, M.H., Ertekin, C., Menges, H.O., Haciseferogullari, H., Evrendilek, F., 2010. Assessing monthly average solar radiation models: a comparative case study in Turkey. *Environmental Monitoring and Assessment*, 175, 251-277.
- Trabea, A., Shaltout, M.M. 2000. Correlation of global solar radiation with meteorological parameters over Egypt. *Renewable Energy*, 21(2):297-308.
- Yao, W., Li, Z., Wang, Y., Jiang, F., Hu, L., 2014. Evaluation of global solar radiation models for Shanghai, China. *Energy Conversion and Management*, 84, 597-612.
- Zeng, J., Qiao, W. 2013. Short-term solar power prediction using a support vector machine. *Renewable Energy* 52:118-127.
- Mohammadi, K., Shamsirband, S., Tong, C.W., Alam, K.A., Petkovic, D., 2015. Potential of adaptive neuro-fuzzy system for prediction of daily global solar radiation by day of the year. *Energy Conversion and Management*, 93, 406-413.
- Mohammadi, B., Emamgholizadeh, S. (2017) Using principal component analysis to inputs the effective rainfall estimates based on entries to help support vector machine and artificial neural network. *Journal of Rainwater Catchment Systems*. 2017; 4 (4) ,67-75. (In Farsi)
- Mohammadi, B., Aghashariatmadari, Z., Moazenzadeh, R. 2018. Determination of Input Variables to Estimate Solar Radiation Using Entropy Theory and Principal Component Analysis. *Iranian Journal of Soil and Water Research*, 50(3): 626-639. (In Farsi)
- Montgomery, D. C., Peck, E. A., Vining, G. G. 2012. *Introduction to linear regression analysis*, 5th Edition, Hoboken, NJ: Wiley.
- Mubiru, J., Banda, E.J.K.B., D'Ujanga, F., Senyonga, T., 2007. Assessing the performance of global solar radiation empirical formulations in Kumpala, Uganda. *Theoretical and Applied Climatology*, 87, 179-184.
- Noori, R., Karbassi, A. and M. Sabahi. 2010. Evaluation of PCA and gamma test techniques on ANN operation for weekly solid waste prediction. *Journal of Environmental Management*, 91(3): 767-771. 20.
- Okundamiya, M.S., Emagbetere, J.O., Ogujor, E.A. 2016. Evaluation of various global solar radiation models for Nigeria. *Energy Conversion and Management*, 13(5):505-512.
- Ozgoren, M., Bilgili, M., Sahin, B., 2012. Estimation of global solar radiation using ANN over Turkey. *Expert Systems with Applications*, 39, 5043-5051.
- Piri, J., Kisi, O., 2015. Modelling solar radiation reached to the Earth using ANFIS, NN-ARX, and empirical models (case studies: Zahedan and Bojnurd stations). *Journal of Atmospheric and Solar-Terrestrial Physics*, 123, 39-47.
- Rahimikhoob A. 2010. Estimating global solar radiation using artificial neural network and air temperature data in a semi-arid environment. *Renewable Energy*, 35, 2131-2135.
- Sabzi-Parvar, A., Aqelpour, P., Warshawian, V., 2019. Comparison of multivariate linear regression performance and artificial intelligence models in



Prediction of solar radiation intensity in Yazd station by using regression model based on principal components (PCR)

S. Soltani-Gerdefaramarzi^{1*}

Received: 19/07/2022

Accepted: 13/04/2023

Abstract

Solar radiation reaching the earth's surface is one of the most important variables used in solar energy projects, hydrological modeling, evaporation and transpiration, meteorology and agriculture. In this research, the functionality of principal component analysis and linear multivariate regression was investigated in predicting the amount of solar radiation in Yazd station between 2010 to 2014 and 2015 to 2021. Variables of average temperature, minimum temperature, maximum temperature, hours of sunshine, relative humidity and solar radiation are received from the Meteorological Organization on a daily basis and variables of extraterrestrial solar radiation, relative distance from the earth to the sun, solar inclination angle and maximum hours of sunshine are calculated with the existing relations and was chosen as input of Principal component analysis method. The correlation of these parameters with the amount of solar radiation shows the presence of a significant positive relationship at the level of 1% between the amount of solar radiation and extraterrestrial radiation, the ratio of sunny hours, the average temperature and the relative distance from the earth to the sun, and with the solar inclination angle at the level of 5% and was observed a significant negative correlation at the level of 1% with relative humidity values. The results of the analysis of the main components showed that the first two components can be selected as the main components according to the eigenvalues of the parameters and the percentage of variance. The correlation coefficient of the first and second components with solar radiation was 0.893 and -0.168, respectively, and as a result, the first component has a higher correlation with the amount of solar radiation than the second component. The rotated value of loading in the first component shows that the minimum temperature, maximum temperature, average temperature and relative humidity are more correlated with solar radiation and the correlation of other factors (extraterrestrial radiation, solar inclination angle and relative distance from the earth to the sun and the ratio of sunny hours) is low in this component. Investigating the relationship between solar radiation and the values of the main components obtained from PC1 and PC2 using multivariate linear regression showed that the effect of meteorological parameters on solar radiation is significant ($p < 0.001$) and its R^2 is was determined as 0.89%. The evaluation criteria of PCR and MLR methods show the ability of PCA method to estimate solar radiation

Keywords: Multivariate regression, Principal components analysis, Solar radiation, VIF, Yazd



¹ Associated Professor, Department of Water Engineering and Sciences, Faculty of Agriculture and Natural Resources, Ardakan University

(*Corresponding Author Email Address: ssoltani@ardakan.ac.ir)

نحوه ارجاع مقاله:

سلطانی گردفامرزی، س. ۱۴۰۲. پیش‌بینی تابش خورشیدی در ایستگاه یزد با بکارگیری مدل رگرسیونی مبتنی بر مؤلفه‌های اصلی (PCR). نشریه

هواشناسی کشاورزی، ۱۱(۱): ۶-۱۶. DOI: 10.22125/agmj.2023.352446.1140

Soltani-Gerdefaramarzi, S. 2023. Prediction of solar radiation intensity in Yazd station by using regression model based on principal components (PCR). Journal of Agricultural Meteorology, 11(1): 6-16. DOI: 10.22125/agmj.2023.352446.1140