

پیش بینی تبخیر و تعرق گیاه مرجع (ET_0) با استفاده از روش آنالیز مؤلفه های اصلی (PCA) و توسعه مدل رگرسیونی خطی چندگانه (MLR-PCA) (مطالعه موردی: ایستگاه مشهد)

نونا شیخ‌الاسلامی^{۱*} - بیژن قهرمان^۲ - ابوالفضل مساعدی^۳ - کامران داوری^۴ - مهدی مهاجرپور^۵

تاریخ دریافت: ۱۳۹۲/۶/۲۰

تاریخ پذیرش: ۱۳۹۳/۱/۲۶

چکیده

تبخیر و تعرق یکی از مهمترین پارامترهایی است که دانستن آن جهت برآورد نیاز آبی گیاه و طراحی سیستم های آبیاری ضروری است، این پدیده متأثر از پارامترهای اقلیمی زیادی می باشد. در این پژوهش به بررسی تأثیر نسبی متغیرهای موثر در این پدیده و نیز برآورد تبخیر و تعرق مرجع با استفاده از آنالیز مؤلفه های اصلی و تحلیل عاملی پرداخته شده است. به این منظور از داده های ایستگاه سینوپتیک مشهد با اقلیم نیمه خشک واقع در استان خراسان رضوی در مقیاس روزانه و در فاصله سال های ۲۰۰۵-۱۹۹۱ استفاده گردید و تأثیر نسبی پارامترهای دما (T) (حداکثر، متوسط و حداقل)، رطوبت نسبی (RH)، ساعات آفتابی (R_s) و سرعت باد در ارتفاع دو متری از سطح زمین (U_2) بر این پدیده مورد بررسی قرار گرفت. با استفاده از مؤلفه های اصلی استخراج شده، مدل رگرسیونی خطی چندگانه برای تخمین تبخیر و تعرق ارائه شد. آماره t برای مقدار ثابت و برای هر یک از مؤلفه های اصلی تعیین شد که همه ضرایب در سطح ۵٪ معنی دار بودند. در شرایط اقلیمی مشهد، همبستگی مؤلفه های اول (PC_1) و دوم (PC_2) با تبخیر و تعرق به ترتیب دارای ضرایب ۰/۶۹۴ و ۰/۵۵۶ بود، مؤلفه اول نسبت به مؤلفه دوم همبستگی بیشتری با تبخیر و تعرق داشت و دمای حداقل، دمای متوسط، دمای حداکثر و رطوبت نسبی برای برآورد تبخیر و تعرق به ترتیب از اهمیت بیشتری نسبت به سایر متغیرها (ساعات آفتابی و سرعت باد) برخوردار بودند. مقادیر ضریب تبیین روش های MLR-PCA و MLR برابر با ۰/۹۰۳ و ۰/۸۹۷ به دست آمد که بیانگر توانایی روش PCA است.

واژه های کلیدی: آنالیز چند متغیره، ایستگاه مشهد، تحلیل عاملی، مدل سازی

مقدمه

داده های اقلیمی و همچنین برای ارزیابی سایر روش ها پیشنهاد شده است (۷). یکی از مشکلات این روش کثرت اطلاعات مورد نیاز برای محاسبه تبخیر و تعرق است. از آنجایی که داده های مربوط به همه متغیرهای موثر بر تبخیر و تعرق در اکثر ایستگاه ها وجود نداشته و یا ثبت نشده است لذا باید روی متغیرهای موجود آنالیز مناسبی انجام داد تا بتوان تأثیر نسبی آنها را روی این پدیده بررسی کرد. بنابراین می توان از روش های آماری چند متغیره در این خصوص استفاده نمود. روش آنالیز مؤلفه های اصلی (PCA) و تحلیل عاملی (FA) از متداول ترین روش های آماری چند متغیره می باشند که روابط مشاهده شده بین چندین متغیر را نشان می دهند و اهمیت نسبی متغیرهای موثر بر یک پدیده را بیان می کنند. مطالعاتی که در گذشته توسط روش تحلیل عاملی در زمینه های هیدرولوژی، هیدرومتئولوژی، منابع آب و محیط زیست انجام شده است بیانگر

تبخیر و تعرق یکی از مهمترین پارامترهایی است که دانستن آن جهت برآورد آب مصرفی گیاه و طراحی سیستم های آبیاری ضروری است. این پدیده متأثر از متغیرهای زیادی می باشد که می توان به بارندگی، سرعت باد، ساعات آفتابی، فشار سطح، رطوبت نسبی، سطح ابرناکی، خصوصیات خاک و گیاه و کیفیت آب اشاره کرد. از بین روش های تجربی متعدد ارائه شده برای محاسبه تبخیر و تعرق گیاه مرجع، در سال ۱۹۹۰ از سوی کمیسیون بین المللی آبیاری و زهکشی و سازمان خوار و بار جهانی روش فائو پنمن مانیتیت به عنوان تنها روش استاندارد برای محاسبه تبخیر و تعرق گیاه مرجع از روی

۱، ۲، ۴ و ۵ - به ترتیب دانش آموخته کارشناسی ارشد، استاد، دانشیار و دانشجوی دکتری گروه مهندسی آب، دانشکده کشاورزی، دانشگاه فردوسی مشهد
(Email: Nona.sh88@gmail.com) * - نویسنده مسئول

۳ - استاد دانشکده منابع طبیعی و محیط زیست، دانشگاه فردوسی مشهد

شده بین سال های ۱۹۹۱ تا ۲۰۰۵ میلادی که فاقد مشکلات فوق بودند، استفاده گردید. در این تحقیق روش فائو پنمن - ماتیت به عنوان روش استاندارد برای محاسبه تبخیر و تعرق و مقایسه نتایج با آن در نظر گرفته شد. آموزش مدل بر اساس ۷۰ درصد از سری داده ها انجام گرفت و برای مرحله آزمون از ۳۰ درصد باقی مانده داده ها استفاده شد. ورود داده ها به مدل به صورت خام باعث کاهش سرعت و دقت محاسبات می شود، برای حل این مشکل داده های خام بایستی نرمال گردند. برای نرمال کردن داده ها از رابطه ۱ استفاده شد که داده های ورودی را بین ۰/۱ تا ۰/۹ استاندارد می کند:

$$X_{standard} = \frac{(x_i - x_{min})}{(x_{max} - x_{min})} \quad (1)$$

که در آن $X_{standard}$: آمار مشاهده ای استاندارد شده، x_i : مشاهده i ام، x_{min} : کوچکترین مشاهده و x_{max} : بزرگترین مشاهده در بین داده ها می باشد.

روش آنالیز مؤلفه های اصلی

در روش آنالیز مؤلفه های اصلی P متغیر اصلی همبسته به p مؤلفه غیرهمبسته یا متعامد تبدیل می شوند. با اعمال PCA متغیرهای ورودی اصلی به متغیرهای جدید که بدون همبستگی می باشند، تبدیل می شوند. مؤلفه های ایجاد شده ترکیبی خطی از متغیرهای اصلی می باشند (۱۳). اگر در P متغیر اصلی تنها مقداری همبستگی معنی دار وجود داشته باشد انجام روش PCA می تواند مفید باشد. به همین دلیل اولین گام برای اجرای این روش تعیین و تشخیص مناسب بودن داده ها می باشد. روش هایی وجود دارد که محقق از طریق آنها می تواند این مطلب را بیازماید. از جمله این روش ها می توان به محاسبه فاکتور KMO^1 اشاره کرد. مقدار این عامل همواره بین صفر و یک در نوسان است در صورتی که KMO کمتر از ۰/۵۰ باشد، داده ها برای روش تحلیل مؤلفه های اصلی مناسب نخواهند بود و اگر مقدار آن بین ۰/۵۰ تا ۰/۶۹ باشد داده ها متوسط و اگر مقدار این شاخص، بزرگتر از ۰/۷۰ باشد همبستگی های موجود در بین داده ها برای تحلیل مؤلفه های اصلی مناسب خواهند بود. این عامل به کمک ضرایب همبستگی ساده^۲ و جزئی^۳ طبق رابطه ۲ محاسبه می شود:

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

1- Kaiser-Meyer-Olkin

2- Simple correlation coefficients

3- Partial correlation coefficients

کارایی این روش می باشد. والیس (۱۹) بیان کرد که روش تحلیل عاملی ابزار عددی مناسبی برای گزینش متغیرهایی است که تاثیر زیادی در یک پدیده دارند. تاکنون از روش آنالیز مؤلفه های اصلی برای اهداف متفاوتی نظیر ارزیابی و پایش ایستگاه های هواشناسی جهت تعیین ایستگاه های اصلی و فرعی (۲)، بررسی خصوصیات شیمیایی آب های زیرزمینی و داده های مربوط به شوری خاک (۹) و تغییرات مکانی و زمانی آلودگی آب های زیرزمینی (۱۶ و ۱۸) استفاده شده است. در بحث اقلیم شناسی نیز با استفاده از روش PCA در پهنه بندی بارندگی زمستانه (۳) و روند تغییر پارامترهای اقلیمی در یک منطقه (۲۰) مطالعاتی انجام پذیرفته است. همچنین در بررسی داده های مربوط به کیفیت آب (۱۴ و ۲۱)، ارزیابی میزان آلودگی رودخانه قره سو (۱۰)، تعیین لایه های خشکسالی در رودخانه های ترکیه (۸)، تعیین مهم ترین پارامترهای تاثیر گذار بر میزان آلودگی رودخانه هراز (۱۷)، شناسایی داده های پرت و مشکوک در حوضه سیلابی دریاچه ارومیه (۵) و نیز پیش بینی دمای آب سطح خلیج فارس (۴)، از روش آنالیز مؤلفه های اصلی استفاده شده است. علاوه بر این برای بررسی داده های بارندگی و رواناب در حوزه اوهایی و بررسی تاثیر نسبی متغیرهای هواشناسی بر تبخیر و تعرق در هند (۱۵) نیز مطالعاتی انجام شده است.

تنها مطالعه ای که برای پیش بینی تبخیر و تعرق مرجع با روش PCA در ترکیب با FA در شرایط مشابه این پژوهش در ایران انجام شده است، مطالعه سیفی و همکاران (۱) در ایستگاه کرمان است. با وجود نزدیکی اقلیم کرمان به مشهد و به دلیل اهمیت داشتن تاثیر نسبی متغیرهای مختلف بر تبخیر و تعرق، بخصوص در شرایط اقلیمی گرم و خشک و لزوم مطالعه بیشتر در این زمینه، این پژوهش به بررسی تاثیر نسبی متغیرهای مختلف بر تبخیر و تعرق در ایستگاه مشهد با شرایط اقلیمی نیمه خشک پرداخته است.

مواد و روش ها

منطقه مورد مطالعه و داده های تحقیق

در این پژوهش از داده های روزانه ایستگاه هواشناسی سینوپتیک مشهد در استان خراسان رضوی استفاده شد. این ایستگاه دارای عرض جغرافیایی ۳۶ درجه و ۱۶ دقیقه شمالی و طول جغرافیایی ۵۹ درجه و ۲۸ دقیقه شرقی می باشد. داده های آماری مورد استفاده در این پژوهش از سازمان هواشناسی استان خراسان رضوی دریافت شد. این داده ها شامل دمای حداکثر (T_{max}, C°)، دمای حداقل (C°)، T_{max} ، دمای متوسط (T_{max}, C°)، رطوبت نسبی (% RH)، ساعات آفتابی (R_{ss}, hr) و سرعت باد در ارتفاع دو متری از سطح زمین ($U_2, m/s$) می باشند. به دلیل ناقص بودن اطلاعات ثبت شده در برخی سال ها و یا حتی عدم وجود اطلاعات، از داده های روزانه ثبت

که در آن: PC_i : i امین مؤلفه مورد نظر، a_{ij} : i امین بردار ویژه متناظر با P امین مؤلفه X_p : P امین متغیر اصلی می باشد.

با انتخاب چند مؤلفه اصلی اول، سایر مؤلفه ها از محاسبات بعدی حذف می شوند. از نمودار واریزه ای (اسکری پلات) که در آن مقادیر ویژه در مقابل شماره مؤلفه ها رسم می شود، برای تشخیص آستانه حذف استفاده می شود. در این روش، مرز بین مؤلفه های اصلی و غیراصلی محلی است که نمودار میل به خطی شدن به صورت افقی می نماید؛ یعنی محلی که مقادیر ویژه در مقابل تغییر شماره مؤلفه، تغییر چندانی نمایند. به منظور دستیابی به ماتریسی ساده و از نظر تئوریک معنادار و قابل تفسیر از چرخش عامل ها استفاده می شود. عاملها به دور مبدأ چرخش داده می شوند تا موقعیت جدیدی کسب کنند. روش چرخش واریماکس به دنبال مختصر کردن و ساده سازی ستون های ماتریس عاملی است. در نتیجه، روش واریماکس استفاده وسیعتری در بین محققین دارد و یکی از انواع چرخش های عمودی است (۱۲).

مدل رگرسیون خطی چند متغیره

مدل رگرسیون خطی چند گانه در فرم ماتریسی را می توان به صورت رابطه ۶ نشان داد:

$$Y = X\beta + \varepsilon \quad (6)$$

که در آن: β : ماتریس ضرایب رگرسیون، ε : ماتریس خطای برازش، Y : ماتریس پاسخ (متغیر وابسته) و X : ماتریس متغیرهای مستقل می باشد. با حل رابطه ۶ بر حسب β خواهیم داشت:

$$\beta = (X'X)^{-1} (X'Y) \quad (7)$$

که در آن X' : ترانزپوز ماتریس X است. برای محاسبه وارون $(X'X)$ ، لازم است قبل از ساخت مدل رگرسیونی، همبستگی بین متغیرهای مستقل را از بین برد. در این خصوص، روش مناسب استفاده از تحلیل مؤلفه های اصلی روی متغیرهای مستقل ورودی به مدل است. ورود متغیرها به مدل رگرسیونی به صورت مرحله ای، از مهم ترین متغیر تا کم اهمیت ترین آنها، صورت می گیرد. معیار میزان اهمیت متغیر در مدل، مقدار سطح معنی دار یا آماره t متناظر با آن در جدول آزمون معنی داری متغیرها می باشد. جهت مدل سازی، مقادیر به دست آمده از PC_1 و PC_2 (رابطه ۵) به عنوان متغیرهای مستقل (برآورد کننده) و مقادیر تبخیر و تعرق به روش فائو پنمن - ماتیت به عنوان متغیر وابسته وارد مدل رگرسیون خطی چندگانه گردیدند.

در این پژوهش برای انجام تجزیه تحلیل مؤلفه های اصلی و مدل سازی تبخیر و تعرق توسط مدل رگرسیونی، از نرم افزار SPSS

که در آن a_{ij} : ضریب همبستگی ساده بین متغیرهای i و j ، a_{ij} : ضریب همبستگی جزئی بین متغیرهای i و j برابر با تعداد متغیرها می باشد. با توجه به رابطه ۲ مقادیر بزرگتر KMO در نتیجه کوچک بودن ضرایب همبستگی جزئی بوده و منجر به افزایش دقت محاسبات با استفاده از روش PCA می شود. در صورتی که مقدار KMO بزرگتر از ۰/۵۰ به دست آید، امکان اجرای روش PCA روی داده های اصلی وجود دارد. دامنه نوسان P متغیر اصلی متفاوت است. از این رو بر اساس رابطه ۳ داده های ورودی به نحوی استاندارد می شوند که دارای میانگین صفر و انحراف معیار یک باشند.

$$Z = \frac{x_{ij} - \mu_i}{\sigma_i} \quad (3)$$

که در آن: Z : معادل استاندارد شده مشاهدات، x_{ij} : داده ورودی برای متغیر i ام و مشاهده j ام، μ_i : میانگین متغیر i ام و σ_i : مقادیر انحراف معیار برای متغیر i ام می باشد.

از ماتریس واریانس - کوواریانس برای نشان دادن میزان تغییرات در نمونه و میزان همبستگی P متغیر با هم استفاده می شود. این ماتریس، ماتریسی متقارن است که عضوهای روی قطر اصلی این ماتریس، واریانس متغیرهای ورودی و بقیه درایه های این ماتریس، کوواریانس بین متغیرهای ورودی است، چون برای تشکیل این ماتریس از داده های استاندارد شده استفاده شده است؛ به همین دلیل این ماتریس، معادل ماتریس همبستگی بین متغیرهای ورودی است بنابراین درایه های روی قطر اصلی در تمامی این ماتریس ها یک می باشد و درایه های غیر قطری کوواریانس بین متغیرهای ورودی است. هر مقدار ویژه با اطلاعات مربوط به آن (بردارهای ویژه) ویژگی های یک مؤلفه را نشان می دهد. هر مؤلفه نیز درصدی از اطلاعاتی که توسط متغیرهای اولیه بیان می شود را در بر می گیرد. هر چه کمیت عددی مقادیر ویژه بزرگتر باشد، بیانگر این است که مؤلفه متناظر با آن نیز درصد بیشتری از اطلاعات متغیرهای اولیه را شامل می گردد. برای به دست آوردن این مقادیر بایستی رابطه ۴ حل شود:

$$|R - \lambda I_p| = 0 \quad (4)$$

که I_p : ماتریس واحد با بعد $P \times P$ ، R : ماتریس واریانس - کوواریانس یا ماتریس همبستگی، λ : مقادیر ویژه و $|...|$: علامت دترمینان ماتریس می باشد. معادله فوق یک مسئله کلاسیک مقادیر ویژه است. λ ریشه ی مشخصه و a بردار ویژه ی R می باشند. بنابراین می توان P مقدار ویژه مرتب شده را بصورت $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p$ بدست آورد، به طوری که مجموع مقادیر ویژه برابر با P باشد و دترمینان ماتریس فوق برابر با صفر گردد. هر مؤلفه، ترکیبی خطی از متغیرهای اولیه می باشد و با رابطه ۵ مشخص می شود:

$$PC_i = a_{i1}X_1 + a_{i2}X_2 + \dots + a_{ip}X_p \quad (5)$$

ثابت مانده است. به عبارت دیگر، می توان گفت که ظاهراً پارامتر سرعت باد در ایستگاه مشهد نسبت به دیگر پارامترها دارای تاثیر کمتری بر تبخیر و تعرق است. به منظور بررسی دقیق تر تاثیر نسبی این پارامتر و سایر پارامترها بر روی تبخیر و تعرق از روش آنالیز مؤلفه های اصلی استفاده شده است.

پیش پردازش متغیرهای ورودی به مدل رگرسیونی با استفاده از PCA

مقدار پارامتر KMO در ایستگاه مشهد برابر با ۰/۶۸۶ می باشد که نشان دهنده امکان اجرای روش PCA روی داده هاست. پس از استاندارد کردن متغیرهای ورودی، ماتریس همبستگی (R) از مرتبه ۶ (معادل با تعداد متغیرهای ورودی) برای متغیرهای اولیه تشکیل می گردد که نتایج آن در جدول ۲ بیان شده است. این ماتریس نشان می دهد که تمامی متغیرهای مورد استفاده در این پژوهش با یکدیگر همبستگی دارند به این صورت که همبستگی متغیرهای دما (T_{max} ، T_{min} ، T_{mean}) با یکدیگر و با متغیر ساعات آفتابی (R_s)، زیاد و مثبت و همبستگی آنها با رطوبت نسبی (RH)، زیاد و در جهت منفی می باشد. همبستگی سرعت باد (U_2) با سایر متغیرها کم بوده و همبستگی ساعات آفتابی (R_s) با رطوبت نسبی (RH)، زیاد و در جهت منفی است. سیفی و همکاران (۱) نیز در پژوهش خود در کرمان با اقلیمی مشابه مشهد؛ نشان دادند که همبستگی ساعات آفتابی (R_s) و دما (T) با رطوبت نسبی (RH) زیاد و در جهت منفی می باشد.

با حل دستگاه معادله ۴ تعداد ۶ مقدار ویژه به دست می آید. جدول ۳ نشان دهنده مقادیر ویژه است. مطابق این جدول مقدار اولین مؤلفه که همان مقدار ویژه اول (λ_1) است برابر با ۴/۲۷۹ می باشد که ۷۱/۳۲ درصد از کل واریانس داده ها را در بر می گیرد. دومین مقدار ویژه نیز ۱۶/۳ درصد از کل واریانس داده ها را شامل می گردد و این دو مؤلفه مجموعاً ۸۷/۶۲ درصد از پراکندگی داده های اصلی را بیان می کنند. بنابراین در ایستگاه مشهد، می توان دو مؤلفه اول را به عنوان مؤلفه اصلی انتخاب کرد. انتخاب دو مؤلفه مناسب است زیرا در نمودار واریزه ای (شکل ۱) نیز تغییرات مقادیر ویژه با شماره مؤلفه ها برای مؤلفه های بزرگ تر از ۲ عملاً تشکیل خطی افقی را داده اند. سیفی و همکاران (۱) نیز در مطالعه خود در ایستگاه کرمان دو مؤلفه اول را به عنوان مؤلفه های اصلی در نظر گرفتند. این دو مؤلفه هم ۸۰ درصد از پراکندگی داده ها را بیان می کردند. نوری قیداری (۵) در مطالعه ای در مورد منطقه سیلابی دریاچه ارومیه، سه مؤلفه اول که ۸۷/۵ درصد از پراکندگی داده ها را بیان می کردند، به عنوان مؤلفه های اصلی در نظر گرفت.

معیارهای ارزیابی عملکرد

به منظور مقایسه مدل ها با یکدیگر از پارامترهای سنجش آماری زیر استفاده گردید. این پارامترها عبارتند از: میانگین مربعات خطا (RMSE^۱، رابطه ۸)، خطای مطلق میانگین (MAE^۲، رابطه ۹)، میانگین نسبی خطای مطلق (MARE^۳، رابطه ۱۰) و ضریب تبیین (R^2 ، رابطه ۱۱)، که بجز ضریب تبیین، هر چقدر مقادیر محاسبه شده این پارامترها کمتر باشد نشان دهنده دقت بالاتر و برتری مدل مربوط به آن می باشد. در مورد ضریب تبیین نیز، محدوده تغییرات آن بین صفر و یک بوده و مقادیر بالاتر آن تطابق بهتر داده های مشاهده ای و برآورد شده را نشان می دهد. گاهی اوقات نتایج بدست آمده از چهار معیار اول متفاوت هستند که در این حالت شاخص جدید خطای نقطه ایده آل (IPE^۴) به صورت ترکیبی از چهار شاخص دیگر در انتخاب بهترین مدل کمک خواهد کرد. مقدار این شاخص بین صفر تا یک متغیر است و هر چه به صفر نزدیکتر باشد بهتر است (۶).

$$RMSE = \left(\frac{1}{N} \sum_{i=1}^N (q_i - O_i)^2 \right)^{0.5} \quad (8)$$

$$MAE = \frac{1}{N} \sum_{i=1}^N |q_i - O_i| \quad (9)$$

$$MARE = \frac{1}{N} \sum_{i=1}^N \frac{|q_i - O_i|}{O_i} \quad (10)$$

$$R^2 = 1 - \frac{SSE}{SST} \quad (11)$$

$$IPE = \left[\left(\frac{RMSE_{ij}}{\max RMSE_{ij}} \right)^2 + \left(\frac{MAE_{ij}}{\max MAE_{ij}} \right)^2 + \left(\frac{MBE_{ij}}{\max MBE_{ij}} \right)^2 + \left(\frac{R_{ij}-1}{\sqrt{\max R_{ij}}} \right)^2 \right]^{0.25} \quad (12)$$

در این روابط، N: تعداد نمونه ها، q_i : مقادیر پیش بینی شده تبخیر و تعرق با استفاده از مدل، O_i : مقادیر محاسبه شده تبخیر و تعرق با استفاده از فرمول فائو- پمن - مانتیث است.

نتایج و بحث

میانگین سالانه پارامترهای آماری در ایستگاه مشهد در جدول ۱ نشان داده شده است. همانطور که مشاهده می گردد، بجز پارامتر سرعت باد، میانگین سالانه دیگر پارامترها تغییر چندانی نداشته است. همچنین مقدار تبخیر و تعرق مرجع نیز در طول این سه سال تقریباً

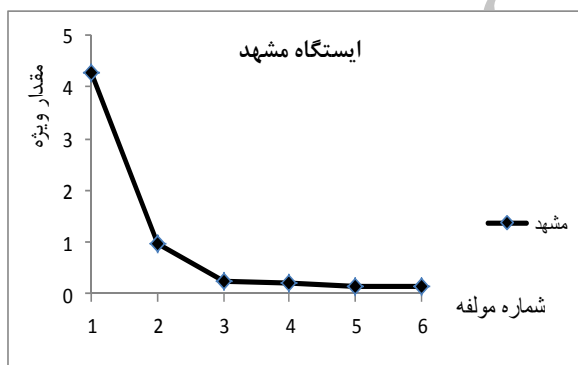
- 1- Root Mean Square Error
- 2- Mean Absolute Error
- 3- Mean Absolute Relative Error
- 4- Ideal Point Error

جدول ۱- میانگین سالیانه پارامترهای آماری در ایستگاه مشهد

ET ₀ (mm/d)	U ₂ (m/s)	R _s (hr)	RH(%)	T _{mean} (°C)	T _{min} (°C)	T _{max} (°C)	سال
۵/۰۸	۵/۸۹	۷/۶۴	۵۸/۴۹	۱۵/۲۲	۸/۷۰	۲۱/۷۳	۲۰۰۳
۵/۰۳	۴/۵۱	۸/۱۷	۵۷/۹۶	۱۵/۹۷	۹/۴۷	۲۲/۴۷	۲۰۰۴
۵/۰۴	۳/۷۶	۸/۲۴	۵۸/۸۷	۱۵/۸۶	۹/۷۷	۲۱/۹۵	۲۰۰۵

جدول ۲- ماتریس مقارن همبستگی بین متغیرهای مورد استفاده در ایستگاه سینوپتیک مشهد

U ₂	R _s	RH	T _{mean}	T _{min}	T _{max}	
۰/۲۴۳	-۰/۷۴۵	-۰/۸۳۶	-۰/۹۸۴	۰/۹۱۷	۱/۰۰	T _{max}
۰/۳۳۸	-۰/۵۸۲	۰/۶۶۹	-۰/۹۷۳	۱/۰۰	۰/۹۱۷	T _{min}
۰/۲۹۰	-۰/۶۸۸	-۰/۷۷۹	۱/۰۰	۰/۹۷۳	۰/۹۸۴	T _{mean}
-۰/۱۷۱	-۰/۷۷۸	۱/۰۰	-۰/۷۷۹	-۰/۶۶۹	-۰/۸۳۶	RH
۰/۱۰۲	۱/۰۰	-۰/۷۷۸	۰/۶۸۸	۰/۵۸۲	۰/۷۴۵	R _s
۱/۰۰	-۰/۱۰۲	-۰/۱۷۱	-۰/۲۹۰	۰/۳۳۸	۰/۲۴۳	U ₂



شکل ۱- نمودار واریزه ای (scree plot) در ایستگاه سینوپتیک مشهد

جدول ۳- مشخصات مقادیر ویژه حاصل از پارامترها در ایستگاه سینوپتیک مشهد

مقادیر ویژه اولیه			مؤلفه
درصد تجمعی	درصد از واریانس	کل	
۷۱/۳۲	۷۱/۳۲	۴/۲۷۹	۱
۸۷/۶۲	۱۶/۳۰	-۰/۹۷۸	۲
۹۵/۸۳	۸/۲۱۲	-۰/۴۹۳	۳
۹۹/۳۳۵	۳/۵۰۳	-۰/۲۱۰	۴
۹۹/۹۹	۰/۶۶۴	-۰/۰۴۰	۵
۱۰۰	-۰/۰۰۱	۳/۸۳×۱۰ ^{-۵}	۶

در جدول ۴ مقادیر بردارهای ویژه مربوط به ایستگاه مشهد آورده شده است که ضرایب هر مؤلفه را برای محاسبه آنها تعیین می کند. برای تشکیل مؤلفه اول بایستی مقادیر متغیر پارامترها را در بردارهای ویژه مربوط به هر متغیر ضرب نموده و حاصل را با هم جمع کرد. یعنی دمای حداکثر را در ۰/۹۸، مقادیر متغیر دمای حداقل را در

و به همین ترتیب مقادیر بقیه متغیرها را در ضرایب مربوط ضرب کرد؛ در نتیجه مؤلفه هایی حاصل می شود که می توان از آنها به جای متغیرهای اولیه به عنوان ورودی به مدل رگرسیون استفاده نمود (روابط ۱۳ و ۱۴):

$$PC_1 = (0.98 \times T_{max}) + (0.919 \times T_{min}) + (0.973 \times T_{mean}) + (-$$

که وابستگی عوامل بارگذاری با مؤلفه های اصلی آشکارتر شده و تشخیص اینکه تغییرات هر یک از مؤلفه های اصلی متاثر از کدام عوامل بارگذاری می باشد آسان تر گردد. جدول ۵ بیانگر مشخصات مؤلفه های اصلی پس از چرخش واریماکس در ایستگاه مشهد می باشد. با توجه به جدول فوق در ایستگاه مشهد، ضریب همبستگی مؤلفه های اول و دوم با تبخیر و تعرق به ترتیب ۰/۶۹۴ و ۰/۵۵۶ می باشد. بنابراین مؤلفه اول نسبت به مؤلفه دوم همبستگی بیشتری با ET_0 دارد و در نتیجه متغیرهایی که در این مؤلفه از نظر مقدار عددی سهم بیشتری دارند (بارگذاری بزرگتری دارند) همبستگی بیشتری را با ET_0 خواهند داشت. مقادیر دوران داده شده بارگذاری، در مؤلفه اول نشان می دهد که دمای حداقل، دمای متوسط، دمای حداکثر و رطوبت نسبی به ترتیب همبستگی بیشتری را با ET_0 دارند و همبستگی بقیه عوامل (ساعات آفتابی و سرعت باد) در این مؤلفه، کم می باشد. به گونه ای که همبستگی پارامتر دما در جهت مثبت و همبستگی رطوبت نسبی در جهت منفی است. در مقابل در مؤلفه دوم بیشترین بارگذاری مربوط به متغیر ساعات آفتابی و سرعت باد می باشد.

$$PC_2 = (0.877 \times RH) + (0.808 \times R_s) + (0.324 \times U_2) \quad (13)$$

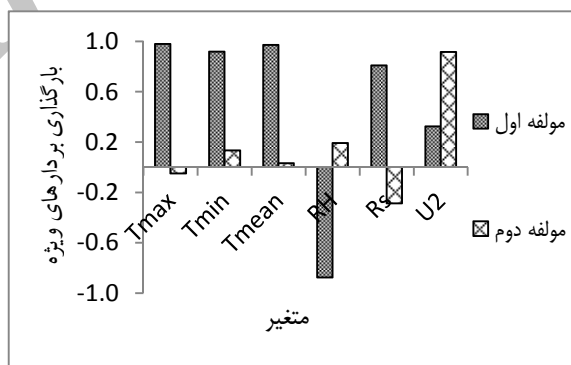
$$PC_2 = (-0.051 \times T_{max}) + (0.134 \times T_{min}) + (0.031 \times T_{mean}) + (0.192 \times RH) + (-0.288 \times R_s) + (0.915 \times U_2) \quad (14)$$

شکل ۲ بردارهای ویژه دو مؤلفه اول در ایستگاه مشهد را نشان می دهد که با استفاده از جدول ۳ تشکیل شده است. با توجه به این شکل، در اولین مؤلفه بیشترین مقدار قدر مطلق بردار ویژه به ترتیب مربوط به متغیرهای دمای حداکثر، دمای متوسط، دمای حداقل، رطوبت نسبی و ساعات آفتابی و در مؤلفه دوم بیشترین مقدار مربوط به سرعت باد می باشد. در مطالعه موردی در ایستگاه کرمان نیز در اولین مؤلفه بیشترین مقدار مربوط به متغیرهای دمای حداکثر و دمای حداقل بوده است و در مؤلفه دوم بیشترین مقدار مربوط به متغیر دمای نقطه شبنم گزارش شده است (۱). نتایج این بررسی تا حد زیادی با نتایج تحقیق حاضر در ایستگاه مشهد مشابهت دارد که احتمالاً نزدیکی شرایط اقلیمی دو استان در ایجاد این شرایط نقش دارد.

چون در تشکیل هر مؤلفه از تمام متغیرهای اولیه استفاده می شود، بنابراین از روش چرخش واریماکس برای تفسیر ساده تر استفاده شده و همبستگی دو مؤلفه اول نسبت به متغیرهای مورد استفاده و تبخیر و تعرق بررسی می گردد. این چرخش سبب می گردد

جدول ۴- بردارهای ویژه مربوط به ایستگاه سینوپتیک مشهد

پارامترها	۱	۲	۳	۴	۵	۶
T_{max}	۰/۹۸۰	-۰/۰۵۱	۰/۱۲۵	-۰/۰۱۹	۰/۱۴۵	۰/۰۰۳
T_{min}	۰/۹۱۹	۰/۱۳۴	۰/۳۴۱	-۰/۰۷۹	-۰/۱۲۴	۰/۰۰۲
T_{mean}	۰/۹۷۳	-۰/۰۳۱	۰/۲۲۴	-۰/۰۲۵	۰/۰۲۷	-۰/۰۰۵
RH	-۰/۸۷۷	۰/۱۹۲	۰/۲۶۲	-۰/۳۵۰	۰/۰۴۹	$3/427 \times 10^{-6}$
R_s	۰/۸۰۸	-۰/۲۸۸	-۰/۴۲۸	۰/۲۸۳	-۰/۰۱۸	$1/613 \times 10^{-6}$
U_2	۰/۳۲۴	۰/۹۱۵	-۰/۲۴۱	-۰/۰۰۲	۰/۰۰۹	$9/133 \times 10^{-7}$



شکل ۲- مقادیر بردارهای ویژه مربوط به بارگذاری مؤلفه های اول و دوم در ایستگاه مشهد

موهان (۱۵) نیز در مطالعه خود روی نواحی گرم و خشک بیان کرد که دمای حداکثر، دمای حداقل و رطوبت نسبی در مؤلفه اول دارای بارگذاری بزرگتری هستند و در نواحی مرطوب رطوبت نسبی و سرعت باد دارای بارگذاری بزرگتری هستند. فنگ و همکاران (۱۱) نیز در پژوهش خود روی پارامترهای موثر در تبخیر و تعرق در نواحی گرم و خشک به این نتیجه رسیدند که پارامترهای دما و ساعات آفتابی دارای تاثیر مثبت و سرعت باد و رطوبت نسبی دارای تاثیر منفی روی تغییرات تبخیر و تعرق می باشند. در مطالعه موردی انجام شده در ایستگاه کرمان (۱) نیز همبستگی متغیر سرعت باد در مؤلفه اول با تبخیر و تعرق کم و با دمای حداکثر و رطوبت نسبی زیاد می باشد که نتایج پژوهش ما را تایید می نماید. با توجه به نزدیکی شرایط اقلیمی دو ایستگاه کرمان و مطالعه حاضر (ایستگاه مشهد)، به نظر می رسد که پارامترهای دما و رطوبت نسبی در برآورد تبخیر و تعرق در نواحی گرم و خشک دارای اهمیت بیشتری نسبت به دیگر پارامترها باشند و تغییرات تبخیر و تعرق در این نواحی بیشتر متاثر از تغییرات این پارامترها می باشد. با وجود اهمیت کمتر پارامترهای ساعات آفتابی و سرعت باد در مؤلفه اول، به دلیل بارگذاری بزرگتر این پارامترها در مؤلفه دوم، امکان حذف این پارامترها در محاسبات مربوط به تبخیر و تعرق وجود ندارد.

موهان (۱۵) نیز در مطالعه خود روی نواحی گرم و خشک بیان کرد که دمای حداکثر، دمای حداقل و رطوبت نسبی در مؤلفه اول دارای بارگذاری بزرگتری هستند و در نواحی مرطوب رطوبت نسبی و سرعت باد دارای بارگذاری بزرگتری هستند. فنگ و همکاران (۱۱) نیز در پژوهش خود روی پارامترهای موثر در تبخیر و تعرق در نواحی گرم و خشک به این نتیجه رسیدند که پارامترهای دما و ساعات آفتابی دارای تاثیر مثبت و سرعت باد و رطوبت نسبی دارای تاثیر منفی روی تغییرات تبخیر و تعرق می باشند. در مطالعه موردی انجام شده در ایستگاه کرمان (۱) نیز همبستگی متغیر سرعت باد در مؤلفه اول با تبخیر و تعرق کم و با دمای حداکثر و رطوبت نسبی زیاد می باشد که نتایج پژوهش ما را تایید می نماید. با توجه به نزدیکی شرایط اقلیمی دو ایستگاه کرمان و مطالعه حاضر (ایستگاه مشهد)، به نظر می رسد که پارامترهای دما و رطوبت نسبی در برآورد تبخیر و تعرق در نواحی گرم و خشک دارای اهمیت بیشتری نسبت به دیگر پارامترها باشند و تغییرات تبخیر و تعرق در این نواحی بیشتر متاثر از تغییرات این پارامترها می باشد. با وجود اهمیت کمتر پارامترهای ساعات آفتابی و سرعت باد در مؤلفه اول، به دلیل بارگذاری بزرگتر این پارامترها در مؤلفه دوم، امکان حذف این پارامترها در محاسبات مربوط به تبخیر و تعرق وجود ندارد.

ارزیابی کارایی PCA

به منظور مقایسه نتایج روش MLR-PCA با مقادیر حاصل از روش MLR تمامی متغیرهای مورد استفاده در این پژوهش وارد مدل رگرسیونی گردید و معادله رگرسیونی خطی چند متغیره حاصل گردید (معادله ۱۶). با توجه به این معادله، آماره t برای مقدار ثابت و برای هر یک از متغیرهای دمای حداکثر، دمای حداقل، دمای متوسط، رطوبت نسبی، ساعات آفتابی و سرعت باد در سطح ۵٪ معنی دار می باشد:

$$ET_0 = -0.115 - 0.544 T_{max} + 0.511 T_{min} + 0.558 T_{mean} - 0.238 RH + 0.148 R_s + 0.435 U_2 \quad (16)$$

جدول ۶- آماره های خطاستنچی مدل های (MLR و MLR-PCA)

مرحله آموزش		مرحله آزمون		معیار
MLR	MLR-PCA	MLR	MLR-PCA	
۰/۰۸۳	۰/۰۷۷	۰/۰۶۶	۰/۰۶۵	RMSE
۰/۰۶۴	۰/۰۶۱	۰/۰۵۱	۰/۰۵۰	MAE
۰/۵۸۱	۰/۵۶۲	۰/۶۹۸	۰/۶۴۶	MARE
۰/۹۵۳	۰/۹۶۱	۰/۹۵۰	۰/۹۵۰	R ²
۰/۵۰۱	۰/۳۶۷	۰/۵۱۱	۰/۳۵۷	IPE

در جدول ۶ معیارهای آماری محاسبه شده جهت مقایسه مدل ها در ایستگاه مشهد در دو مرحله آموزش و آزمون بیان شده است، بر اساس نتایج این جداول برتری اندکی بین روش MLR-PCA نسبت به MLR در ایستگاه مشهد وجود دارد ولی به طور کلی مدل آنالیز مؤلفه های اصلی نسبت به مدل رگرسیون خطی چند متغیره تفاوت معنی داری ندارد. سیفی و همکاران (۱) نیز در مطالعه خود بر روی تبخیر و تعرق در ایستگاه کرمان برتری اندکی بین دو روش MLR-PCA و MLR نشان دادند، آنها در مطالعه خود هفت پارامتر هواشناسی را مورد بررسی قرار داده بودند، اما فتاحی (۱۰) در مطالعه ای روی رودخانه قره سو با در نظر گرفتن ۲۸ پارامتر موثر بر

جدول ۵- مؤلفه های اصلی پس از چرخش واریماکس در ایستگاه

نام پارامتر	شماره مؤلفه	
	۱	۲
T _{max}	۰/۸۲۵	۰/۵۳۴
T _{min}	۰/۹۳۱	۰/۲۷۹
T _{mean}	۰/۸۹۰	۰/۴۳۱
RH	-۰/۴۹۰	-۰/۷۹۳
R _s	۰/۳۲۴	۰/۹۰۱
U ₂	۰/۱۵۸	۰/۹۳۳
ET ₀	۰/۶۹۴	۰/۵۵۶

ساخت مدل رگرسیونی

جهت مدل سازی، مقادیر مؤلفه های به دست آمده PC₁ و PC₂ به عنوان متغیرهای مستقل (برآورد کننده) و مقادیر تبخیر و تعرق به روش فائو - پنمن به عنوان متغیر وابسته وارد مدل رگرسیون خطی چندگانه گردیدند و رابطه ۱۵ به دست آمد:

$$ET_0 = -0.193 + 0.233 PC_1 + 0.344 PC_2 \quad (15)$$

در این روش، آماره t برای مقدار ثابت و برای هر یک از مؤلفه های PC₁ و PC₂ به ترتیب برابر با ۲۲۴/۰۲۷، -۷۲/۲۰۵ و

گردید. به این منظور داده های ایستگاه سینوپتیک مشهد واقع در استان خراسان رضوی مورد استفاده قرار گرفت و شش پارامتر دمای حداکثر، دمای حداقل، دمای متوسط، رطوبت نسبی، ساعات آفتابی و سرعت باد در ارتفاع ۲ متری از سطح زمین برای محاسبات در نظر گرفته شدند. نتایج حاصل نشان داد که، دو مؤلفه اول که مجموعاً ۸۷/۶۲ درصد از پراکندگی داده های اصلی را بیان می کردند به عنوان مؤلفه های اصلی در نظر گرفته شدند و بر مبنای آنها مدل رگرسیون خطی ارائه گردید. در مرحله بعد مدل رگرسیونی با استفاده از تمامی متغیرها تعیین شد. مقایسه بین پارامترهای آماری در هر دو مدل نشان داد که بین دو مدل MLR-PCA و MLR تفاوت معنی داری وجود ندارد.

کیفیت آب، برتری مطلق روش آنالیز مؤلفه های اصلی را نسبت به رگرسیون نشان داد.

به نظر می رسد که روش آنالیز مؤلفه های اصلی در مواردی که تعداد متغیرها زیاد باشد بر رگرسیون خطی برتر است و در حالت کم بودن تعداد متغیرها (مانند پژوهش حاضر) احتمالاً تعداد کم پارامترهای مورد استفاده نمی تواند برتری مدل MLR-PCA را نسبت به MLR به خوبی نشان دهد.

نتیجه گیری

با توجه به اهمیت ET_0 در محاسبات آبیاری و نیز لزوم استفاده از روش های نوین در برآورد تبخیر و تعرق مرجع، در این پژوهش از روش آنالیز مؤلفه های اصلی برای تخمین تبخیر و تعرق استفاده

منابع

- ۱- سیفی ا، میرلطیفی س.م. و ریاحی ح. ۱۳۸۹. توسعه مدل ترکیبی رگرسیون چندگانه-تحلیل مؤلفه ها و عامل های اصلی (MLR-PCA) در پیش بینی تبخیر و تعرق مرجع (مطالعه موردی: ایستگاه کرمان). آب و خاک، ۲۴ (۶): ۱۱۹۶ - ۱۱۸۶.
- ۲- سیفی ا، میرلطیفی س.م. و ریاحی ح. ۱۳۹۰. ارزیابی و پایش شبکه ایستگاههای هواشناسی به روش تحلیل مؤلفه های اصلی و تحلیل عاملی. (مطالعه موردی: ایستگاه کرمان). آبیاری و زهکشی ایران، ۱ (۵): ۴۲ - ۳۰.
- ۳- ناظم السادات س.م.ج، بیگی ب. و امین س. ۱۳۸۲. پهنه بندی بارندگی زمستانه استانهای بوشهر، فارس و کهگیلویه و بویراحمد با استفاده از روش تحلیل مؤلفه های اصلی. علوم و فنون کشاورزی و منابع طبیعی، ۱ (۷): ۷۱ - ۶۱.
- ۴- ناظم السادات س.م.ج. و شیروانی ا. ۱۳۸۴. پیش بینی دمای سطح آب خلیج فارس با استفاده از رگرسیون چند گانه و تحلیل مؤلفه های اصلی. علوم و فنون کشاورزی و منابع طبیعی، ۹ (۳): ۱ تا ۱۰.
- ۵- نوری قیداری م.ح. ۱۳۸۹. شناسایی داده های پرت در آنالیز منطقه ای سیلاب به روش آنالیز مؤلفه های اصلی. پنجمین کنگره ملی مهندسی عمران، ۱۴ تا ۱۶ اردیبهشت، دانشگاه فردوسی مشهد، ۵۳۱ - ۵۲۴.
- ۶- محمدی مقدم س، مساعدی ا، جنگجو م. و مصداقی م. ۱۳۹۲. مدل سازی روابط پارندگی و خشکسالی با تولید مراتع ندوشن یزد. دومین کنفرانس بین المللی مدل سازی گیاه، خاک، آب و هوا، ۱۸ تا ۱۹ اردیبهشت، دانشگاه کرمان، ۶۲-۵۱.
- ۷- وزیري ژ، سلامت ع، انتصاری م، مسچی م، حیدری م. و دهقانی سانچ ح. ۱۳۸۷. دستورالعمل محاسبه آب مورد نیاز گیاهان. کمیته ملی آبیاری و زهکشی ایران، تهران.
- 8- Demirel M. 2007. Performing k-means analysis to drought principal components of Turkish rivers. Rosenstiel School of Marine and Atmospheric Sciences, Division of Meteorology and Physical Oceanography . Miami Florida.
- 9- 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.
- 10- Fataei E. 2011. Assessment of surface water quality using principle component analysis and factor analysis. World Journal of Fish and Marine Sciences, 3 (2): 159-166.
- 11- Feng J., Yan D., Li Ch., Yu F., and Zhang Ch. 2013. Assessing the impact of climatic factors on potential evapotranspiration in droughts in North China. Journal of Quaternary International, (Article in press).
- 12- Johnson R.A. and Wichern D.W. 1982. Applied multivariate statistical analysis. Prentice-Hall Inc., Englewood Cliffs, SA, 590 p.
- 13- Liu C.W., Lin K.H. and 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.
- 14- Mazlum N., Ozer A. and Mazlum S. 1999. Interpretation of water quality data by principal components analysis. Tur. Journal of Engineering and Environmental Science, 23: 19-26.
- 15- Mohan S., and Arumugam N. 1996. Relative importance of meteorological variables in evapotranspiration: Factor analysis approach. Journal of Water Resources Management, 10: 1-20.

- 16- Ouyang Y. 2005. Application of principal component and factor analysis to evaluate surface water quality monitoring network. *Journal of Water Resources Research*, 39: 2621-2635.
- 17- Pejman A.H., Nabi Bidhendi G.R., Karbassi A.R., Mehrdadi N., and Esmaeili Bidhendi M. 2009. Evaluation of spatial and seasonal variations in surface water quality using multivariate statistical techniques. *Environment Science and Technology*, 6: 467-476.
- 18- Petersen W. 2001. Process identification by principal component analysis of river water quality data. *Journal of Ecological Modeling*, 138: 193-213.
- 19- Wallis J.R. 1968. Factor analysis in hydrology: An agnostic view. *Journal of Water Resources Research*, 4(3): 521-527.
- 20- Wan K.L. 2009. A new variable for climate change study and implications for the built environment. *Journal of Renewable Energy*, 34: 916-919.
- 21- Zhu C., Jia L., Liang Q., and Peng W. 2011. Application of factor analysis in evaluation of water quality. *International Journal of Advancements in Computing Technology*, 3:4156-4200.

Archive of SID

Estimating Reference Evapotranspiration by Using Principal Component Analysis (PCA) and The Development of a Regression Model (MLR-PCA) (Case Study: Mashhad Station)

N. Sheikholeslami^{1*} - B. Ghahraman² - A. Mosaedi³ - K. Davary⁴ - M. Mohejerpour⁵

Received: 11-09-2013

Accepted: 15-04-2014

Abstract

Evapotranspiration is one of the most important parameters that its understanding is necessary for estimating crop water requirement and design of irrigation systems. This phenomenon is greatly influenced by climatic parameters. In this study, the relative importance of variables affecting this phenomenon was evaluated and the reference evapotranspiration was estimated using principal component analysis and factor analysis. Daily scaled measurements for the period of 1991-2005 were obtained from synoptic stations located in Mashhad Khorasan Razavi province, Iran. Mashhad has a semi-arid climate area. The measurements included the relative influence of temperature (T) (maximum, average and minimum), relative humidity (RH), sunshine hours (Rs), and the wind speed at a height of two meters above the ground (U2). The multiple linear regressions were used to estimate evapotranspiration. T-statistic with a significant level of 5% was used for the main components. The evapotranspiration was correlated more with T (minimum, maximum, and average), and relative humidity as than wind speed or sunshine. PC1 had more effect than PC2 (with coefficients of 0.694 and 0.556, respectively). MLR-PCA and MLR with coefficients of 0.903 and 0.897 (respectively) indicated higher ability for PCA method.

Keywords: Multivariate analysis, Mashhad station, Factor analysis, Modeling

Archive of SID

1,2,4, 5- MSc Graguated, Professor, Associate Professor and PhD Student, Department of Water Engineering, Faculty of Agriculture, Ferdowsi University of Mashhad, Respectively

(* - Corresponding Author Email: Nona.sh88@gmail.com)

3- Professor Faculty of Natural Resources and Environment, Ferdowsi University of Mashhad