Archive of SID

Journal of Water and Soil
Vol. 24, No. 6, Jan-Feb 2011, p. 1186-1196

نشریه آب و خاک
جلد 24، شماره 6، بهمن - اسفند 1389، ص. 1196-1186

توسعه مدل ترکیبی رگرسیون چندگانه- تحلیل مولفه‌ها و عامل‌های اصلی (MLR-PCA) در پیش‌بینی تبخیر-تعرق مرجع

(مطالعه موردی: ایستگاه کرمان)

اکرم سیفی[1]- سید مجید میرلطیفی[2*]- حسین ریاحی[3]

تاریخ دریافت:89/1/22
تاریخ پذیرش:89/7/11

چکیده

تبخیر-تعرق مرجع یکی از پارامترهای مهم در مدیریت آبیاری گیاهان است. تبخیر-تعرق مرجع یک پدیده چندمتغیره و پیچیده می‌باشد که چندین متغیر هیدرولوژیکی آن را تحت تاثیر قرار می‌دهند و معمولا بر مبنای پایگاه داده‌های هواشناسی چندساله با استفاده از مدل‌های نیمه تجربی برآورد می‌شود. اهمیت کاربردی تخمین دقیق تبخیر- تعرق مرجع، پیچیدگی و ناشناخته بودن ریاضیات پدیده، لزوم استفاده از روش‌های جدید داده‌کاوی را نشان می‌دهد. به همین دلیل در این مقاله، امکان برآورد تبخیر-تعرق مرجع با استفاده از مدل ترکیبی رگرسیون چندگانه و تحلیل مولفه‌های اصلی (MLR-PCA) بررسی شد و اهمیت نسبی متغیرهای موثر بر تبخیر-تعرق مرجع با استفاده از تحلیل عاملی مورد ارزیابی قرار گرفت. داده‌های هواشناسی روزانه سال‌های 2005-1996 ایستگاه سینوپتیک کرمان در این تحلیل استفاده شد. دو مولفه PC1 و PC2 که 80 درصد واریانس کل را شرح دادند به عنوان مولفه‌های اصلی و بقیه به عنوان اختلال در نظر گرفته شدند. با استفاده از مولفه‌های اصلی استخراج شده، مدل رگرسیون خطی چندگانه برای تخمین تبخیر-تعرق مرجع ارائه شد. آماره t برای مقدار ثابت و برای هر یک از مولفه‌های اصلی تعیین گردید که طبق نتایج، تمامی ضرایب در سطح 5 درصد معنی‌دار بودند. طبق نتایج، PC1 اهمیت بیشتری نسبت به مولفه دیگر دارد و در مرحله بعدی PC2 دارای اهمیت می‌باشد و بنابراین مقادیر متغیرهای شدت تابش، رطوبت نسبی، ساعات آفتابی، دمای حداقل و دمای حداکثر برای برآورد تبخیر-تعرق از اهمیت بیشتری نسبت به سایر متغیرها برخوردارند. مقادیر ضریب همبستگی روش‌های MLR-PCA و MLR بر اساس مبنای مقایسه‌ای فائو-پنمن-مانتیث در مرحله آزمون، به ترتیب 0/820 و 0/840 بدست آمد که این مقادیر اختلاف چندانی نداشته و بیانگر توانایی روش PCA کاهش تعداد متغیرهای مورد استفاده است.

واژه های کلیدی: تبخیر-تعرق مرجع، فائو- پنمن- مانتیث، رگرسیون چندگانه، تحلیل مولفه اصلی، تحلیل عاملی

مقدمه

تمامی پدیده‌های هیدرولوژیکی به طور طبیعی پدیده‌های تصادفی هستند. این پدیده‌ها را می‌توان با استفاده از تعدادی متغیر تعیین کرد که ممکن است این متغیرها با یکدیگر دارای همبستگی زیادی باشند. تبخیر-تعرق پدیده‌ی هیدرولوژیکی چند بعدی است که تحت تاثیر چندین متغیر عمل می‌کند. به طور کلی، متغیرهایی که در فرآیند تبخیر-تعرق اثرگذارند، بارندگی، سرعت باد، تابش خورشیدی، فشار سطحی، رطوبت، ساعات آفتابی، ابرناکی، همرفت[4] (انتقال افقی گرمای محسوس و رطوبت)، پوشش زمین، خصوصیات خاک و گیاه و مقدار آب می‌باشد (22). برآورد دقیق نیاز آبی گیاهان یکی از راه‌های کاهش تلفات آب در مزارع می‌باشد (9). بنابراین شناخت تاثیر نسبی متغیرها بر تبخیر-تعرق مرجع[5] (ET_o) در رابطه با مدیریت آبیاری بسیار مهم می‌باشد. از آنجایی که اکثر متغیرهای تاثیرگذار در تمامی ایستگاه‌ها اندازه‌گیری نمی‌شود، بنابراین تحلیل داده‌های موجود برای درک اهمیت نسبی این متغیرها لازم است. یکی از روش‌های آماری چندمتغیره[6] که امروزه به طور وسیعی مورد استفاده قرار می‌گیرد، روش

1و2- دانشجوی کارشناسی ارشد و دانشیار گروه آبیاری و زهکشی، دانشکده کشاورزی، دانشگاه تربیت مدرس
(*- نویسنده مسئول: Email:mirlat_m@modares.ac.ir)
3- دانشجوی دکتری سازه‌های آبی، دانشکده کشاورزی، دانشگاه تربیت مدرس

4- Advection
5- Reference Evapotranspiration
6- Multivariate

www.SID.ir

Archive of SID

تحلیل مولفه اصلی[1] (PCA) و تحلیل عامل اصلی[2] (PFA) می‌باشد. این روش بین مجموعه‌ای بزرگ از داده‌های به ظاهر بی‌ارتباط، رابطه خاصی را تحت یک مدل فرضی برقرار می‌کند، سپس هر متغیر را به صورت ترکیب خطی از فاکتورهای پنهان تبدیل می‌کند. هدف اصلی روش تحلیل عاملی، یافتن روش مختصر و به دست آوردن اطلاعات مفید از تعداد زیادی متغیر مورد مشاهده با فراوانی بالاست. در این روش متغیرها به مجموعه‌های کوچکتری از عامل‌ها با کمترین اطلاعات از دست رفته و موثرترین اطلاعات موثر بر پدیده تبدیل می‌شوند. در تحلیل عاملی، متغیرهایی که در یک عامل تعریف می‌شوند، به یکدیگر کاملاً وابسته‌اند و این وابستگی، عامل را بوجود می‌آورد. از طرفی متغیرهای هر عامل هیچ وابستگی به متغیرهای عامل‌های دیگر ندارد (14). در سال‌های اخیر روش‌های آماری چند متغیره در موضوعات مرتبط با منابع آب، هیدرولوژی و محیط زیست به طور گسترده‌ای استفاده شده‌اند. اخیراً، استفاده از روش‌های PCA و PFA برای کاهش تعداد متغیرهای ورودی و تفسیر بهتر نتایج به دست آمده از تحلیل داده‌های کیفیت آب، رواج یافته است (25). از روش PFA برای تعیین الگوهای زمانی و مکانی بارش‌های روزانه (1)، شناخت عوامل موثر بر سیلاب (13)، تعیین عامل‌های مهم اقلیمی موثر بر دبی اوج (3)، ارزیابی کیفیت زمانی و مکانی جریان سطحی (26)، پیش‌بینی مکانی محتوای خاک (27)، پیش‌بینی جریان آب با استفاده از MLR-PCA (24) و مطالعه شبکه ایستگاه-های باران‌سنجی و تعیین ایستگاه‌هایی که می‌توان آنها را حذف کرد (23)، استفاده شده است. همچنین در مطالعات مختلف، روش‌های PCA و PFA در استخراج الگوهای زمانی و مکانی برای متغیرهای مختلف هواشناسی به کار گرفته شدند (21). منطقه‌بندی بارندگی زمستانه در ناحیه جنوب مرکزی ایران (استان‌های فارس، بوشهر و کهگیلویه و بویراحمد) (10)، شناسایی الگوهای زمانی- مکانی و پهنه‌بندی رژیم‌های دمایی ماهانه (6) با استفاده از PCA انجام شده است. با استفاده از داده‌های اقلیمی شامل بارش، نم نسبی، ساعات آفتابی، میانگین دما، میانگین کمینه‌ی دما، میانگین بیشینه‌ی دما و فشار تراز ایستگاه در ایستگاه سنندج و انجام تحلیل مولفه‌های اصلی روی این داده‌ها، تیپ‌های هم‌دید هوا به دست آمد (8). برای استخراج مولفه‌های اصلی و کاهش حجم داده‌ها در ارزیابی امکان پیش‌بینی دمای زمستانه سطح آب خلیج فارس با استفاده از مدل رگرسیون چندگانه (11) و پیش‌بینی میانگین غلظت روزانه مونوکسیدکربن در هوای شهر تهران با استفاده از دو مدل شبکه عصبی مصنوعی و رگرسیون خطی چندمتغیره از روش PCA استفاده شده است (12). در تمامی این موارد استفاده از PCA منجر به بهبود دقت در تحلیل نتایج و داده های آماری شده است و نتایج تحقیقات سابق نشان‌دهنده کارآئی مناسب تکنیک تحلیل مولفه ها و عوامل اصلی در مدلسازی پدیده‌های هیدرولوژیکی و اقلیم شناسی مختلف است. از تحلیل مولفه‌های اصلی برای تشخیص چهره استفاده گردیده است و نتایج نشان داد که این روش برای مجموعه داده‌های بزرگ مناسب‌تر می‌باشد و راندمان محاسباتی بالاتری را ایجاد می‌کند (32). تحلیل مولفه‌های اصلی به عنوان ابزاری برای کنترل فرآیند آماری چندگانه تعیین خطا در مواردی که با حجم زیادی از اطلاعات روبرو هستیم، استفاده گردیده است (18). قبل از طبقه‌بندی طیفی ستارگان با شبکه عصبی غیرخطی از PCA برای پیش فرآیندسازی داده‌ها استفاده گردیده است و بیان شده است که این فن باعث می‌شود تا پایداری شبکه، طبقه‌بندی و همگرایی آن افزایش یابد (31). مطالعات مختلفی برای برآورد ET_o با استفاده از روش‌های شبکه عصبی و رگرسیون خطی چندگانه و مقایسه بین این روش‌ها انجام شده است (19 و 28). علیرغم تحقیقات وسیعی که در زمینه مدلسازی ET_o صورت گرفته است، تاکنون از روش‌های PCA و PFA در ترکیب با روش رگرسیون خطی چندگانه برای برآورد ET_o استفاده نشده است و به همین دلیل در این تحقیق، امکان توسعه مدل ترکیبی رگرسیون چندگانه- تحلیل مولفه‌ها و عامل‌های اصلی در برآورد ET_o، در قالب مطالعه موردی، ایستگاه سینوپتیک کرمان بررسی شد و اهمیت نسبی متغیرهای موثر بر تبخیر-تعرق مرجع مورد ارزیابی قرار گرفت. بدین منظور با توجه به آمار و اطلاعات در دسترس و از اطلاعات روزانه ثبت شده در سازمان هواشناسی کشور در فاصله سال‌های 1996 تا 2005 میلادی استفاده گردید. در ادامه مشخصات منطقه مورد مطالعه و داده‌های تحقیق، روش تحقیق و نتایج و بحث ارائه شده است.

مواد و روش‌ها

منطقه مورد مطالعه و و داده‌های تحقیق

استان کرمان در جنوب شرقی ایران و بین 53 درجه و 26 دقیقه تا 59 درجه و 29 دقیقه طول شرقی و 25 درجه و 55 دقیقه تا 32 درجه عرض شمالی قرار دارد و با 185675 کیلومترمربع وسعت، 11 درصد از خاک کشورمان را به خود اختصاص داده است. بر اساس نتایج تحقیق دین پژوه (17)، که کل ایران را به سه اقلیم بسیار خشک، خشک و نیمه خشک و مرطوب تقسیم بندی کرده است، دسته بندی اقلیمی برای ایستگاه سینوپتیک کرمان انجام گرفت که بر اساس آن، این ایستگاه دارای اقلیم بسیار خشک می‌باشد. در این تحقیق، از آمار و اطلاعات موجود از ایستگاه سینوپتیک کرمان در سازمان هواشناسی کشور، استفاده گردید. به دلیل ناقص بودن اطلاعات روزانه ثبت شده قبل از سال 1996 و یا حتی عدم وجود

1- Principal Component Analysis
2- Principal Factor Analysis

www.SID.ir

Archive of SID

آمار در بعضی سال‌ها از اطلاعات روزانه ثبت شده سازمان هواشناسی کشور بین سال‌های 1996 تا 2005 میلادی استفاده گردید. متغیرهای مورد استفاده در این تحقیق دمای حداقل (°C)، دمای حداکثر (°C)، دمای نقطه شبنم (°C)، رطوبت نسبی (٪)، سرعت باد (m/s)، شدت تابش ($MJ/m^2/day$) و ساعات آفتابی(hr) می‌باشد. سازمان فائو، معادلهٔ فائو-پنمن- مانتیث[1] (F-P-M) را به عنوان روش استاندارد ساخت پایگاه داده‌های ET_o در ایستگاه‌های فاقد اطلاعات لایسیمتری برای ارزیابی و واسنجی معادلات دیگر توصیه کرده است. روش F-P-M دارای پایهٔ فیزیکی قوی می‌باشد و برای تخمین درست ET_o در سرتاسر دنیا استفاده شده است (15). در این تحقیق، روش F-P-M به عنوان روش استاندارد برای برآورد ET_o و مقایسه نتایج در نظر گرفته شد. طبق تعریف استاندارد ارائه شده برای تبخیر-تعرق مرجع، در استفاده از معادله فائو-پنمن-مانتیث باید از داده‌های ایستگاه هواشناسی مرجع استفاده شود. با توجه به اینکه در این تحقیق از آمار ایستگاه سینوپتیک استفاده شده و این ایستگاه در منطقه غیر کشاورزی و بدون پوشش واقع است، لذا از شرایط مرجع، برای برآورد تبخیر-تعرق برخوردار نیست، بنابراین لازم است که قبل از به کارگیری داده‌های این ایستگاه نسبت به شرایط مرجع تصحیح شوند. بدین منظور از روش توصیه شده در دستورالعمل FAO-56 استفاده شد (15). به دلیل آنکه وارد کردن داده‌ها به صورت خام باعث کاهش سرعت و دقت مدل می‌شود، در این تحقیق برای جلوگیری از ایجاد مشکلات عددی، نرمال‌سازی داده‌ها در محدوده‌ی 0/1 تا 0/9 انجام گرفت. آموزش مدل بر اساس 50 درصد از داده-های سری زمانی مورد استفاده انجام گرفت و از 50 درصد بقیه در مرحله آزمون مدل استفاده شد.

تحلیل مولفه‌های اصلی

تحلیل مولفه اصلی و تحلیل عاملی از روش‌های آماری چندمتغیره هستند که می‌توان از آنها برای کاهش پیچیدگی تحلیل متغیرهای اولیه مسئله در مواردی که با حجم زیادی از اطلاعات روبرو هستیم و همچنین برای تفسیر بهتر اطلاعات استفاده نمود (16). با این روش، متغیرهای اولیه به مولفه‌های جدید و مستقل (با ضرایب همبستگی صفر برای هر دو مولفه) تبدیل می‌شوند و سپس از این مولفه‌ها به جای متغیرهای اولیه استفاده می‌گردد. مولفه‌های جدید، ترکیبی خطی از متغیرهای اولیه هستند (20). به علاوه چون در تشکیل مولفه‌ها از تمام متغیرها استفاده می‌گردد، در نتیجه اطلاعات متغیرهای اولیه با کمترین تلفات به وسیله مولفه‌های حاصل ارائه می‌شود و باعث از دست دادن جنبه‌های اطلاعاتی داده‌های اصلی نمی‌شود (22). روش کار برای ایجاد مولفه‌های اصلی و تعیین متغیرهای اصلی به صورت زیر می‌باشد.

الف- محاسبه فاکتور KMO

از آنجایی که روش‌های PCA و PFA مستلزم وجود و قبول فرضیاتی درباره جامعه مورد مطالعه نیست، از روش‌های آماری ناپارامتری می‌باشند که لازم است امکان استفاده از روش‌های مذکور و نتایج به دست آمده از آنها به وسیله عامل KMO[2] یا آزمون بارتلت[3]، مشخص شود. مقدار KMO بین صفر تا یک تغییر می‌کند. این عامل با استفاده از ضرایب همبستگی ساده[4] و جزئی[5] طبق رابطه 1 محاسبه می‌شود. در رابطه 1، r_{ij} و a_{ij} ضرایب همبستگی ساده و جزئی بین متغیرهای i و j است. با توجه به رابطه 1 مقادیر بزرگتر KMO مستلزم کوچک بودن ضرایب همبستگی جزئی می‌باشد و بیانگر دقت محاسبات مربوطه، با استفاده از PCA و PFA است.

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

در صورتی که این عامل بزرگتر از 0/5 به دست آید، نشان-دهنده‌ی امکان اجرای این دو روش بر داده‌های اصلی می‌باشد (29).

ب- استاندارد نمودن متغیرهای ورودی: در این مرحله داده‌های ورودی بر اساس فرمول زیر به نحوی استاندارد می‌شوند که دارای میانگین صفر و انحراف معیار یک باشند.

$$Z = \frac{X - \mu}{\sigma} \qquad (2)$$

در این فرمول، Z معادل مقادیر استاندارد شده داده‌ها، X داده‌های ورودی، μ میانگین هر متغیر و σ نیز مقادیر انحراف معیار برای هر متغیر است (12).

ج- محاسبه ماتریس همبستگی (R) برای متغیرهای اولیه: این ماتریس، که ماتریسی متقارن است، میزان تغییرات در نمونه و میزان همبستگی P متغیر را با هم نشان می‌دهد. عضوهای روی قطر اصلی این ماتریس، واریانس متغیرهای ورودی و بقیه درایه‌های این ماتریس، کوواریانس بین متغیرهای ورودی است. چون برای تشکیل این ماتریس از داده‌های استاندارد شده استفاده شده است، به همین دلیل این ماتریس، معادل ماتریس همبستگی بین متغیرهای ورودی است (12).

د- محاسبه مقادیر ویژه (λ) و بردارهای ویژه مربوطه از ماتریس

1- FAO Penman-Monteith

2- Kaiser-Meyer-Olkin
3- Bartlet Test
4- Simple correlation coefficients
5- Partial correlation coefficients

www.SID.ir

Archive of SID

همبستگی: بدین منظور معادله زیر حل می شود:

$$|R - \lambda I_P| = 0 \qquad (3)$$

I_P یک ماتریس واحد با بعد $p \times p$ می‌باشد. بنابراین می‌توان p مقدار ویژه مرتب شده $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_P$ را بدست آورد، بطوری که مجموع مقادیر ویژه برابر با p باشد. هر مقدار ویژه با اطلاعات مربوط به آن (بردارهای ویژه) ویژگی‌های یک مولفه را ارائه می‌دهد. هر مولفه نیز درصدی از اطلاعاتی که توسط متغیرهای اولیه بیان می‌شود را دربرمی‌گیرد و معادل با بخشی از اطلاعات مسئله است که به صورت عدد و رقم در متغیرهای اولیه نهفته است. هر چه کمیت عددی مقادیر ویژه بزرگتر باشد، بیانگر این است که مولفه ایجادی از آن نیز درصد بیشتری از اطلاعات متغیرهای اولیه را دربرمی‌گیرد (12). اولین مولفه بیشترین واریانس و آخرین آن کمترین مقدار واریانس را نشان می‌دهد. انتخاب چند مولفه اول که بیشترین مقدار واریانس را دارند و به عنوان مولفه‌های اصلی شناخته می‌شوند، از اساسی‌ترین اقدامات در تجزیه و تحلیل مولفه‌های اصلی می‌باشد. با انتخاب چند مولفه اصلی اول، سایر مولفه‌ها از محاسبات بعدی حذف می‌شوند و بنابراین باید دقت زیادی در انتخاب آستانه حذف نمود. Screet plot یکی از روش‌های تشخیص آستانه حذف می‌باشد که در آن مقادیر ویژه در مقابل شماره مولفه‌ها رسم می‌شود (10). در این روش، مرز بین مولفه‌های اصلی و غیراصلی محلی است که نمودار میل به خطی شدن می‌نماید یعنی محلی که مقادیر ویژه در مقابل تغییر شماره مولفه، تغییر چندانی ننماید (10).

هـ- اجرای چرخش مناسب روی ماتریس ضرایب مولفه‌ها: چون در تشکیل هر مولفه از تمام متغیرهای اولیه استفاده می‌شود، تفسیر مولفه‌ها مشکل خواهد بود. به این دلیل روش‌هایی برای تفسیر ساده-تر مولفه‌ها به وجود آمده است. این روش‌ها، همان چرخش مولفه‌ها هستند که به دو نوع چرخش عمودی و مایل تقسیم می‌شوند. به دلیل اینکه در روش چرخش عمودی، استقلال بین مولفه‌ها حفظ می‌شود، این نوع چرخش بیشتر مورد استفاده قرار می‌گیرد. در مطالعات علمی بیشتر از چرخش وریماکس استفاده می‌شود که یکی از روش‌های چرخش عمودی است (30). این روش نسبت به بقیه روش‌ها نتایج بهتری را ایجاد می‌کند و به عنوان چرخش استاندارد توصیه می‌گردد. استفاده از چرخش وریماکس برای تفسیر بهتر نتایج، PFA نامیده می‌شود (25). برای اجرای تحلیل مولفه‌های اصلی و تحلیل عامل اصلی از نرم افزارهای آماری مختلفی می‌توان بهره گرفت و در این تحقیق از نرم‌افزار آماری SPSS Statistic 17 برای این منظور استفاده گردید.

مدل رگرسیون خطی چندمتغیره[1] (MLR)

مدل رگرسیون خطی چند گانه در فرم ماتریسی را می‌توان به صورت معادله زیر نشان داد:

$$Y = X\beta + e \qquad (4)$$

در معادله 4، β ماتریس ضرایب رگرسیون، e ماتریس خطای برازش و Y ماتریس پاسخ (متغیر وابسته) و X ماتریس متغیرهای مستقل می‌باشند. با حل معادله 4 بر حسب β خواهیم داشت:

$$\beta = (X'X)^{-1}(X'Y) \qquad (5)$$

که در رابطه 5، X' ترانهاده ماتریس X است. برای محاسبه معکوس $(X'X)$، لازم است متغیرهای مستقل همبستگی زیادی داشته باشند، زیرا در این صورت ماتریس $(X'X)$ را نمی‌توان معکوس کرد و باعث افزایش خطا در اثر گرد کردن داده‌ها و محاسبات می‌شود. برای رفع این مشکل باید قبل از ساخت مدل رگرسیونی، همبستگی بین متغیرهای مستقل را از بین برد. در این خصوص، روش مناسب، استفاده از تحلیل مؤلفه‌های اصلی روی متغیرهای مستقل ورودی به مدل است. در این تحقیق پس از رفع مشکل همبستگی در متغیرهای مستقل، مدلی مناسب با استفاده از روش رگرسیون خطی چندمتغیره برای پیش‌بینی تبخیر-تعرق مرجع توسعه یافته و در محاسبات رگرسیونی از الگوریتم گام به گام[2] استفاده شد. در این روش ورود متغیرها به مدل رگرسیون به صورت مرحله‌ای، از مهم‌ترین متغیر تا کم اهمیت‌ترین آنها، صورت می‌گیرد. معیار میزان اهمیت متغیر در مدل، مقدار سطح معنی‌داری یا آماره t متناظر با آن در جدول آزمون معنی‌داری متغیرها است (2 و 12). بررسی اولیه نشان داد که بین متغیرهای ورودی مورد استفاده در این تحقیق همبستگی معنی‌داری وجود دارد که برای از بین بردن این مشکل، از روش PCA استفاده شد.

معیارهای ارزیابی عملکرد

به منظور مقایسه و ارزیابی عملکرد مدل‌های مورد بررسی از پارامترهای میانگین مربعات خطا[3] (RMSE)، خطای مطلق میانگین[4] (MAE) و ضریب همبستگی (r) استفاده می‌گردد. فرمول‌های ارائه شده برای این پارامترها بصورت زیر می‌باشد:

$$RMSE = \left(\frac{1}{N}\sum_{i=1}^{N}(P_i - O_i)^2\right)^{0.5} \qquad (6)$$

1-Multiple Linear Regression
2-Stepwise
3- Root Mean Square Error
4- Mean Absolute Error

www.SID.ir

Archive of SID

$$MAE = \frac{1}{N}\sum_{i=1}^{N}\left|P_i - O_i\right| \quad (7)$$

در این معادلات، N تعداد نمونه‌ها، P_i معادل مقادیر پیش‌بینی شده ET_o از مدل، O_i مقادیر محاسبه شده ET_o از روش F-P-M می‌باشند. پارامترهای RMSE و MAE هم بعد با پارامترهایی هستند که از آنها بدست آمده‌اند و r بدون بعد است.

پارامترهای ذکر شده هیچ اطلاعی در مورد نحوه توزیع خطا نمی‌دهند. به این منظور برای ارزیابی مدل‌ها از شاخص میانگین قدرمطلق خطای نسبی (MARE)[1] و تحلیل آستانه خطا (TS)[2] نیز استفاده و نمودار پراکندگی خطای مطلق نسبی نیز استفاده می‌گردد. این دو معیار، نه تنها شاخص عملکرد را بصورت جملاتی از مقادیر برآوردشده ارائه می‌دهند بلکه توزیع خطا را نیز نشان می‌دهند. شاخص TS_X برای مقدار x درصد از برآوردها، نشان‌دهنده توزیع خطا در مقادیر برآورد شده برای هر مدل می‌باشد. این شاخص که برحسب درصد تعریف می‌شود برای مقادیر مختلف قدرمطلق خطای نسبی ارائه می‌شود. مقدار TS برای x درصد از برآوردها طبق رابطه 8 به‌دست می‌آید (4):

$$TS_x = \frac{Y_x}{n}\times 100 \quad (8)$$

که در آن Y_x تعداد پیش‌بینی‌شده (از کل تعداد n) برای هر مقدار مطلق خطای نسبی کمتر از x درصد می‌باشد (4).

نتایج و بحث

در ابتدا نتایج روش تحلیل مولفه‌های اصلی ارائه شده و در ادامه نتایج تحلیل عاملی، رگرسیون چندگانه ترکیبی و ارزش نسبی متغیرهای مستقل ارائه شده است.

پیش پردازش متغیرهای ورودی به مدل رگرسیونی با PCA

برای بررسی امکان اجرای آنالیز مولفه های اصلی از آزمون بارتلت استفاده شد. مقدار KMO=0/613 امکان اجرای PCA را تأیید کرد. برای اجرای این روش، پس از استاندارد کردن متغیرهای ورودی ماتریس متقارن همبستگی از مرتبه 7 (معادل با تعداد متغیرهای ورودی) تشکیل شد که نتایج آن در جدول (1) ارائه شده است. با حل دستگاه معادله (3)، 7 مقدار ویژه و به ازای هر مقدار ویژه 7 بردار ویژه، حاصل می‌شود که با استفاده از آنها، مولفه‌های اصلی از متغیرهای اولیه به دست می‌آید. مشخصات هر مولفه در جدول (2) آورده شده است. در جدول (3) مقادیر بردارهای ویژه آمده است که ضرایب هر مولفه را برای محاسبه آنها تعیین می‌کند. همانطور که در جدول (2) نشان داده شده است، مقدار اولین مولفه برابر 3/768 می‌باشد که 53/834 درصد از کل واریانس موجود در سری داده‌ها را توجیه می‌نماید. دومین مقدار ویژه نیز 26/318 درصد از کل واریانس را توجیه می‌نمایند و این دو مولفه حدود 80 درصد کل پراکندگی داده‌های اصلی را بیان می‌کنند. بنابراین، تقریبا می‌توان دو و سه مولفه اول را به عنوان مولفه اصلی قلمداد نمود. در عین حال، انتخاب آستانه حذف فقط بر اساس قضاوت تخمینی دقیق نبوده و لازم است که آزمون‌های دیگری نیز انجام گیرد. هدف از آزمون‌های بعدی، پیدا کردن یک مبنای فیزیکی برای هر یک از مولفه‌های اصلی می‌باشد. برای پیدا نمودن آستانه حذف، تعداد دو، سه و چهار مولفه اصلی (مولفه‌های داوطلب) در نظر گرفته شد و آزمون لازم برای تشخیص تعداد مولفه‌های مطلوب برای نگهداری به عمل آمد. برای این منظور، عوامل بارگذاری[3] مربوط به هر یک از مولفه‌ها به روش وریماکس دوران داده شدند و وابستگی هر یک از مولفه‌های دوران داده شده به ET_o مورد ارزیابی قرار گرفت. در این تحقیق از دوران دو مولفه اول نتایج مطلوبی به دست آمد که نتایج آن در جدول (4) آمده است. با توجه به جدول (4)، ضرایب 0/726 و 0/654 به ترتیب بیانگر همبستگی مولفه اول و دوم با ET_o می‌باشند. بنابراین مولفه اول نسبت به مولفه اصلی دوم همبستگی بیشتری با ET_o دارد و متغیرهایی که در این مولفه بارگذاری بزرگتری دارند همبستگی بیشتری را با ET_o نشان می‌دهند. مقادیر دوران داده شده بارگذاری مولفه اول نشان می‌دهد که متغیرهای شدت تابش، رطوبت نسبی، دمای حداکثر و ساعات آفتابی همبستگی زیادی با تبخیر–تعرق دارند و همبستگی بقیه عوامل (دمای حداقل، دمای نقطه شبنم و سرعت باد) در این مولفه، کم می‌باشد. در مقابل، بردارهای دوران داده شده بارگذاری متناظر با مولفه دوم همبستگی زیادی با متغیرهای دمای نقطه شبنم، سرعت باد و دمای حداقل دارد. در شکل (1) بارگذاری دو مولفه اول نشان داده شده است که با استفاده از بردارهای ویژه از مولفه‌های اول و دوم جدول (3) تشکیل شده است. در اولین مولفه بیشترین بارگذاری‌ها مربوط به متغیرهای دمای حداکثر، دمای حداقل، رطوبت نسبی، شدت تابش و ساعات آفتابی می‌باشد. در مولفه دوم بیشترین بارگذاری مربوط به متغیر دمای نقطه شبنم می‌باشد. با توجه به جدول (3)، برای تشکیل مولفه اول بایستی مقادیر متغیر دمای حداقل (T_{min}) را در ضریب 0/820، مقادیر متغیر دمای حداکثر (T_{max}) را در ضریب 0/942 و به همین ترتیب مقادیر بقیه متغیرها را در ضرایب مربوطه ضرب کرد. بدین صورت مولفه‌هایی حاصل می-

1- Mean Absolute Relative Error
2- Threshold Statistics
3- Loading

www.SID.ir

Archive of SID

شوند که می‌توان از آنها به جای متغیرهای اولیه به عنوان ورودی بـه مدل رگرسیون استفاده نمود. طبق بحث انجام شده، از بردارهای ویژه به دست آمده در جدول (3)، دو مولفه اصلی به صـورت روابـط (9) و (10) استخراج گردید. در ایـن روابـط متغیرهـای T_{min}، T_{max}، T_{dew}، RH_{mean}، U_2، n و R_s به ترتیب بیانگر دمای حداقل، دمـای حـداکثر، دمای نقطه شبنم، رطوبت نسبی، سرعت باد و شدت تابش می‌باشند.

$$PC1 = (0.820 \times T_{min}) + (0.942 \times T_{max}) + (0.343 \times T_{dew}) - (0.779 \times RH_{mean}) + (0.312 \times U_2) + (0.728 \times n) + (0.926 \times R_s) \quad (9)$$

$$PC2 = (0.511 \times T_{min}) + (0.690 \times T_{max}) + (0.842 \times T_{dew}) + (0.469 \times RH_{mean}) + (0.552 \times U_2) - (0.553 \times n) - (0.191 \times R_s) \quad (10)$$

ساخت مدل رگرسیونی

مقادیر به دست آمده از PC1 و PC2 به عنوان متغیرهای مستقل (برآوردکننـده) وارد مـدل رگرسـیون خطـی چندگانـه گردیـد. معادلـه رگرسیون خطی چندگانه با وارد کـردن دو مولفـه اصـلی بـه صـورت معادله (11) به دست آمد.

$$ET_o = -0.288 + 0.310PC1 + 0.214PC2 \quad (11)$$

جدول 1- ماتریس متقارن همبستگی بین متغیرها

متغیرها	دمای حداقل	دمای حداکثر	دمای نقطه شبنم	رطوبت نسبی	سرعت باد	ساعات آفتابی	شدت تابش
دمای حداقل	1/00	0/833	0/696	-0/428	0/455	0/260	0/628
دمای حداکثر	0/833	1/00	0/407	-0/748	0/187	0/565	0/791
دمای نقطه شبنم	0/696	0/407	1/00	0/237	0/317	-0/139	0/197
رطوبت نسبی	-0/428	-0/748	0/237	1/00	-0/910	-0/691	-0/704
سرعت باد	0/455	0/187	0/317	-0/910	1/00	-0/570	0/189
ساعات آفتابی	0/260	0/565	-0/139	-0/691	-0/570	1/00	0/863
شدت تابش	0/628	0/791	0/197	-0/704	0/189	0/863	1/00

جدول 2- مشخصات مولفه های ایجادی از متغیرهای اولیه (ایستگاه‌ها)

مولفه‌ها	مقدار هر مولفه از 7	درصد اطلاعات متغیرهای اولیه	درصد تجمعی اطلاعات متغیرهای اولیه
اول	3/768	53/834	53/834
دوم	1/842	26/318	80/152
سوم	0/784	11/203	91/355
چهارم	0/458	6/540	97/896
پنجم	0/075	1/077	98/973
ششم	0/043	0/614	99/586
هفتم	0/029	0/414	100/00

جدول 3 - ضرایب هر ایستگاه (بردارهای ویژه) جهت تعیین مولفه‌ها

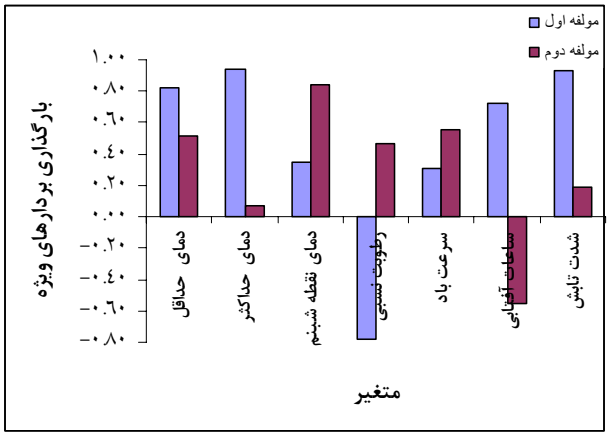
متغیرها	مولفه‌ها						
	اول	دوم	سوم	چهارم	پنجم	ششم	هفتم
دمای حداقل	0/820	0/511	0/073	-0/150	0/170	-0/099	0/020
دمای حداکثر	0/942	0/069	0/158	-0/241	-0/103	0/067	0/102
دمای نقطه شبنم	0/343	0/842	0/364	0/147	0/116	-0/007	-0/074
رطوبت نسبی	-0/779	0/469	0/198	0/346	0/054	0/025	0/101
سرعت باد	0/312	0/552	-0/761	0/128	-0/045	0/007	0/014
ساعات آفتابی	0/728	-0/553	0/039	0/380	-0/070	-0/109	0/031
شدت تابش	0/926	-0/191	0/031	0/274	0/113	0/126	-0/037

www.SID.ir

Archive of SID

جدول 4- مشخصات مولفه‌های اصلی با چرخش وریماکس

متغیرها	مولفه‌ها	
	اول	دوم
دمای حداقل	0/472	0/830
دمای حداکثر	0/793	0/503
دمای نقطه شبنم	-0/100	0/900
رطوبت نسبی	-0/905	0/047
سرعت باد	0/018	0/647
ساعات آفتابی	0/904	-0/139
شدت تابش	0/910	0/279
تبخیر-تعرق مرجع به روش F-P-M	0/726	0/654

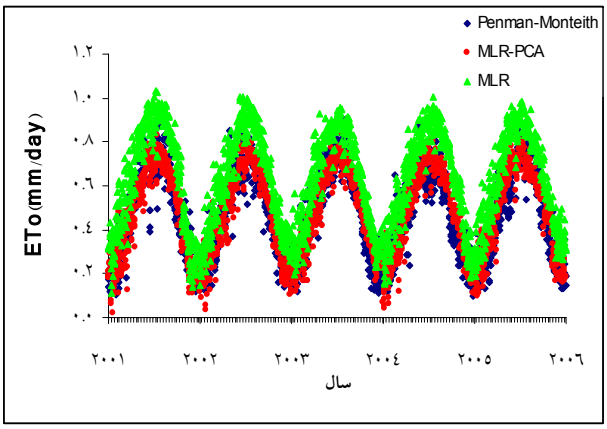


شکل 1- بارگذاری مولفه‌های اول و دوم

در این روش آماره t برای مقدار ثابت و برای هر یک از مولفه-های PC1 و PC2 به ترتیب برابر 50/137-، 151/006 و 44/326 تعیین گردید. با استفاده از جدول t، برای سطح معنی‌دار 5٪ و درجه آزادی 1826، مقدار t=1/645 به دست می‌آید. مقایسه بین آماره‌ها و مقدار به دست آمده از جدول t نشان می‌دهد که تمامی ضرایب در سطح 5٪ معنی‌دار می‌باشند. بزرگی قدر مطلق آماره t نیز اهمیت آن را در مدل MLR-PCA نشان می‌دهد. از این دیدگاه، PC1 اهمیت بیشتری نسبت به مولفه دیگر دارد و در مرحله بعدی، PC2 دارای اهمیت می‌باشند. بنابراین مقادیر متغیرهای شدت تابش، رطوبت نسبی، ساعات آفتابی، دمای حداقل و دمای حداکثر برای برآورد ET_o از اهمیت بیشتری برخوردارند. همچنین به منظور مقایسه نتایج روش MLR-PCA با مقادیر حاصل از روش MLR تمامی متغیرهای مورد استفاده در این تحقیق وارد مدل رگرسیونی و معادله (12) برای آن ارائه گردید. مطابق معادله (12) مقادیر آماره t برای مقدار ثابت و برای هر یک از متغیرهای دمای حداکثر، سرعت باد، شدت تابش، رطوبت نسبی، ساعات آفتابی، دمای حداقل و دمای نقطه شبنم به ترتیب برابر 38/378-، 47/710، 51/542، 51/588، 24/192، 22/190-، 13/845- و 30/953 تعیین گردید. مقایسه بین آماره‌ها و مقدار به دست آمده از جدول نشان می‌دهد که در این معادله نیز تمامی ضرایب در سطح 5٪ معنی‌دار می‌باشند.

$$ET = -0.409 - (0.228 \times T_{min}) + (0.881 \times T_{max}) + (0.508 \times U_2) - (0.291 \times n) + (0.754 \times R_s) + (0.300 \times T_{dew}) + (0.245 \times RH_{mean}) \quad (12)$$

در شکل (2) نمودار سری زمانی مقادیر ET_o برآوردشده با استفاده از مدل MLR-PCA در مقابل مقادیر F-P-M و MLR در دوره مورد مطالعه نشان داده شده است.



شکل 2- نمودار سری زمانی مقادیر برآورد مدل‌ها

در جدول (5) شاخص‌های ضریب همبستگی (r)، مجذور میانگین مربعات خطا (RMSE) و میانگین خطای مطلق (MAE) برای مدل‌های استفاده شده در مرحله آموزشی و آزمایش ارائه گردیده است. همانطور که در جدول (5) نشان داده شده است، ضریب همبستگی مدل MLR-PCA در هر دو مرحله آموزش و آزمون تقریباً به میزان 1٪ کمتر از مدل MLR می‌باشد که این نتیجه بیانگر توانایی مدل PCA در کاهش متغیرها با کمترین مقدار اطلاعات از دست رفته می‌باشد. در حقیقت در مدل MLR-PCA با کاهش متغیرهای ورودی دقت مدل تغیرمحسوسی ننموده است و می توان با متغیرهای کمتر مدلسازی را با دقت مطلوب انجام داد. در شکل (3) نمودار پراکندگی و خط رگرسیونی بین مقادیر ET_o روزانه به دست آمده از روش رگرسیون خطی چندگانه با ورودی دو مولفه در مقابل مقادیر ET_o محاسبه شده از روش F-P-M در مرحله آزمون رسم شده است. مقدار R^2 حاکی از آن است که در مجموع حدود 82 درصد از مقدار واریانس ET_o به وسیله دو مولفه اصلی تعریف می‌شود و این مدل

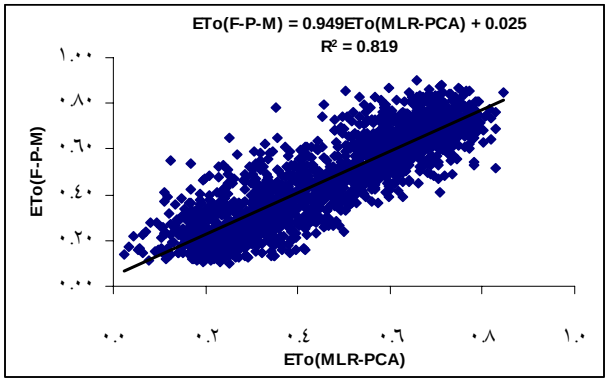
www.SID.ir

Archive of SID

خطای کمتری را نسبت به MLR تولید می‌کند.

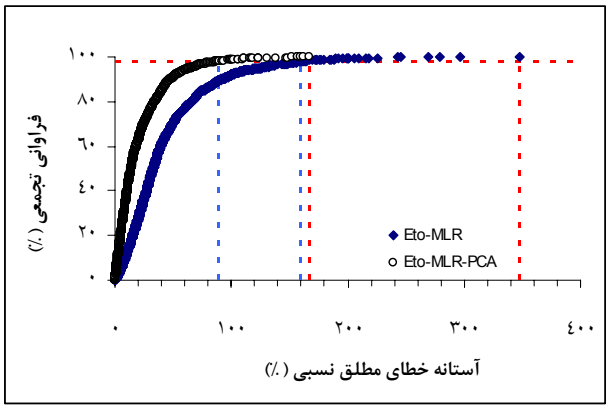
جدول 5- نتایج ارزیابی مدل‌ها

معیار	مرحله آموزش		مرحله آزمون	
	MLR	MLR-PCA	MLR	MLR-PCA
r	0/986	0/974	0/916	0/905
RMSE	0/164	0/055	0/183	0/101
MAE	0/156	0/044	0/160	0/079



شکل 3- نمودار پراکندگی مقادیر برآورد مدل ترکیبی MLR-PCA و F-P-M

در شکل (4) نمودار توزیع خطای پیش‌بینی رابطه (8) آمده است. طبق این شکل مشاهده می شود که در 98 درصد موارد خطای برآورد مدل‌های MLR-PCA و MLR به ترتیب کمتر از 88 و 159 درصد است و فقط برای 2 درصد موارد است که این خطا بیش از 100 درصد است. بنابراین مشخص است که مدل MLR-PCA خطای کمتری را ایجاد کرده و نیز توزیع هموار خطای مدل MLR-PCA بیانگر مطلوبیت آن در برآوردها است.



شکل 4- توزیع خطای برآورد مدل‌های MLR-PCA و MLR

تعیین متغیرهای اصلی

پس از تشکیل ماتریس ضرایب مولفه، اقدام به تعیین پارامترهای اصلی با استفاده از روش‌های PCA و PFA شده است. در روش مورد استفاده، با استفاده از چرخش وریماکس روی ماتریس ضرایب مولفه، بردارهای ویژه به دست آمدند (جدول (6)). در این روش پارامترهای اصلی، پارامترهایی هستند که حداقل یکی از ضرایب آنها که برای تشکیل عامل مربوطه استفاده می‌شود، دارای مقدار نسبتا بالایی باشد. تعیین مقدار این ضریب، به شرایط تحقیق و وسعت منطقه مورد مطالعه و پارامترهای مورد استفاده، بستگی دارد. هر چه منطقه مورد مطالعه دارای وسعت زیادتری باشد، مقادیر کمتری از این ضریب را در نظر گرفت. ولی برای مسائلی که ساده و کوچک باشند، معمولا مقادیر بالاتری برای این ضریب انتخاب می‌شود. در این تحقیق به دلیل وسعت منطقه مورد مطالعه و تعداد پارامترهای کم، این معیار معادل 0/65انتخاب شد (25). با توجه به معیار 0/65 و جدول 6 مشخص می‌شود که برای تمام پارامترها حداقل یک ضریب بالاتر از ضریب 0/65وجود دارد. بنابراین هیچ کدام از پارامترها در این ایستگاه فرعی نمی‌باشند و اندازه‌گیری تمامی آنها لازم است.

جدول 6- مقادیر بردارهای ویژه در روش آنالیز فاکتور اصلی

متغیرها	مولفه‌ها						
	اول	دوم	سوم	چهارم	پنجم	ششم	هفتم
دمای حداقل	0/215	0/729	0/485	0/287	0/322	0/029	0/011
دمای حداکثر	0/420	0/482	0/742	0/049	0/043	0/026	0/187
دمای نقطه شبنم	-0/038	0/984	0/079	0/142	-0/052	0/013	0/011
رطوبت نسبی	-0/466	0/160	0/863	-0/067	-0/024	0/002	0/086
سرعت باد	0/002	0/187	0/063	0/980	0/020	0/008	0/002
ساعات آفتابی	0/984	-0/072	0/280	-0/063	-0/017	-0/112	0/002
شدت تابش	0/845	0/251	0/398	0/114	0/101	0/204	0/022

www.SID.ir

Archive of SID

نتیجه گیری

با توجه به اهمیت برآورد تبخیر-تعرق مرجع (ET_o) در برنامه-ریزی‌های توسعه آبیاری، در این مقاله، امکان برآورد ET_o با استفاده از مدل ترکیبی رگرسیون خطی چندگانه- تحلیل مولفه‌های اصلی (MLR-PCA) و اهمیت نسبی متغیرهای متفاوت بر ET_o مورد ارزیابی قرار گرفت. بدین منظور از داده‌های ایستگاه سینوپتیک کرمان در فاصله سال‌های 1996-2005 و 7 متغیر هواشناسی (متغیرهای مورد استفاده در این تحقیق دمای حداقل (°C)، دمای حداکثر (°C)، دمای نقطه شبنم (°C)، رطوبت نسبی (٪)، سرعت باد (m/s)، شدت تابش ($MJ/m^2/day$) و ساعات آفتابی (hr)) استفاده گردید. با توجه به مشکل همبستگی بین متغیرهای اولیه، در این تحقیق برای اولین بار از روش تحلیل مولفه‌های اصلی به عنوان پیش‌پردازش داده‌های ورودی به مدل تحلیل مولفه های اصلی استفاده شد. با استفاده از آزمون بارتلت پیش فرض‌های استفاده از تحلیل مولفه های اصلی بررسی شد و پس از تائید این آزمون تحلیل مولفه‌های اصلی صورت گرفت. طبق نتایج دو مولفه اصلی که 80 درصد از اطلاعات متغیرهای اولیه را شامل می‌شود، از داده‌های اولیه استخراج گردید و بر مبنای آنها مدل رگرسیون خطی چندگانه ترکیبی ارئه شد. نتایج حاکی از اهمیت زیاد متغیرهای شدت تابش، رطوبت نسبی، دمای حداکثر و ساعات آفتابی بر ET_o در مولفه اول بود که بیشترین تاثیر را در فرآیند ET_o دارند. این دو مولفه به عنوان متغیرهای مستقل وارد مدل رگرسیونی خطی شدند. مشخص گردید که این مولفه‌ها در مجموع 82 درصد از کل واریانس ET_o را توجیه می‌کند و مجذور میانگین مربعات خطای آن 0/101 است. همچنین در این تحقیق اقدام به تعیین متغیرهای اصلی و فرعی شد. با توجه به ضریب 0/65 تمامی متغیرهای مورد ارزیابی، اصلی بودند و نیاز به اندازه‌گیری همه آنها وجود دارد. مدل ترکیبی MLR-PCA ارائه شده در این تحقیق دقت مطلوبی در برآورد ET_o داشته و قابلیت توسعه برای سایر مناطق کشور را داراست. مقایسه نتایج مدل MLR با MLR-PCA نشان دهنده این است که با استفاده از تکنیک تحلیل مولفه‌های اصلی می‌توان به خوبی عوامل اختلال را استخراج نموده و از تحلیل‌ها حذف نمود و سپس مدل پیشگو را با استفاده از متغیرهای غیراختلال و مولفه اصلی توسعه داد. کاهش متغیرهای ورودی تاثیر چندانی بر نتاج مدل پیشگو نداشته و کارائی روش پیشنهادی را تائید نموده است. در ادامه تحقیق حاضر توصیه می‌شود که رهیافت ارائه شده در تحقیق حاضر برای ایستگاه‌های تحقیقاتی که در آنها داده-های لایسیمتری وجود دارد استفاده شود و دقت رویکرد پیشنهادی با استفاده از داده‌های واقعی ارزیابی شود.

منابع

1- جهانبخش اصل س.، و ذوالفقاری ح. 1381. بررسی الگوهای سینوپتیک روزانه در غرب ایران. فصلنامه تحقیقات جغرافیایی. شماره 63، 234-258.

2- خداشناس س.ر.، قهرمان ب.، داوری ک.، و ناظریان ح. 1387. ارائه مدل‌های رگرسیونی چند متغیره برآورد رسوب درحوضه‌های شمال استان خراسان. مجله آب و خاک. شماره 2، صفحات 164-150.

3- روحانی ح.، محسنی ساروی م.، و ملکیان ا. 1383. تعیین مهمترین فاکتورهای اقلیمی و مورفومتری موثر بر دبی اوج و ارائه مدل رگرسیون در شرق و شمال ایران. مجله علوم کشاورزی و منابع طبیعی. شماره 11، صفحات 99-119.

4- ریاحی ح.، و ایوبزاده س.ع. 1387. تخمین ضریب پراکندگی طولی آلودگی با استفاده از سیستم استنتاج فازی- عصبی تطابقی. مجله علمی و پژوهشی آب و فاضلاب. شماره 67، صفحات 46-34.

5- سازمان هواشناسی کشور، www.weather.ir.

6- غیور ح.، و منتظری م. 1383. پهنه‌بندی رژیم‌های دمایی ایران با مولفه‌های مبنا و تحلیل خوشه‌ای. مجله جغرافیا و توسعه. صفحات34-21.

7- فرشادفر ع. 1384. اصول و روشهای آماری چند متغیره. چاپ دوم. انتشارات طاق بستان. 734 صفحه.

8- مسعودیان س.ا.، و محمدی ب. 1386. شناسایی تیپ‌های همدید هوای ایستگاه سنندج (طی سالهای 73-1343). مجله جغرافیا و توسعه ناحیه‌ای. شماره 7.

9- موسوی بایگی س.م.، عرفانیان م.، و سرمد م. 1388. استفاده از حداقل داده‌های هواشناسی برای برآورد تبخیر و تعرق گیاه مرجع و ارائه ضرایب اصلاحی (مطالعه موردی: استان خراسان رضوی). مجله آب و خاک. شماره 1، صفحات100-91.

10- ناظم السادات س.م.ج.، بیگی ب.، و امین س. 1382. پهنه‌بندی بارندگی زمستانه استان‌های بوشهر، فارس و کهکیلویه و بویراحمد با استفاده

www.SID.ir

Archive of SID

از روش تحلیل مولفه‌های اصلی. مجله علوم و فنون کشاورزی و منابع طبیعی. شماره 1، صفحات 71-60.

11-ناظم السادات س.م.ج.، و شیروانی ا. 1384. پیش‌بینی دمای سطح آب خلیج فارس با استفاده از رگرسیون چندگانه و تحلیل مولفه‌های اصلی. علوم و فنون کشاورزی و منابع طبیعی. شماره 3، صفحات 10-1.

12-نوری ر.ا.، اشرفی خ.، و اژدرپور ا. 1387. مقایسه کاربرد روش های شبکه عصبی مصنوعی و رگرسیون خطی چندمتغیره بر اساس تحلیل مولفه های اصلی برای پیش بینی غلظت میانگین روزانه کربن مونوکسید: بررسی موردی شهر تهران. مجله فیزیک زمین و فضا. شماره 34، صفحات 152-135.

13-وفاخواه م. 1378. شناخت عوامل موثر در سیلاب به منظور مهار آنها با استفاده از تجزیه و تحلیل عاملی در رودخانه حوزه قره چای. پژوهش و سازندگی. شماره 45، صفحات72-75.

14-یزدان خواه س. 1387. اهمیت نسبی متغیرهای هواشناسی در برآورد تبخیر-تعرق مرجع در اقلیم های مختلف. پایان نامه کارشناسی ارشد آبیاری و زهکشی، دانشکده کشاورزی، دانشگاه تربیت مدرس، 110 ص.

15- Allen R.G., Pereira L.S., Raes D. and Smith M. 1998. Crop evapotranspiration guidelines for computing crop water requirements, FAO Irrigation and Drainage, Paper 56.

16- Camdevyren H., Demyr N., Kanik A.. and Keskyn S. 2005. Use of principal component scores in multiple linear regression models for prediction of Chlorophyll-a in reservoirs, Ecological Modelling, 181: 581-589.

17- Dinpashoh Y. 2006. Study of reference crop evapotranspiration in I.R. of Iran, Agricultural Water Management, 84: 123-129.

18- Kano M., Hasebe S., Hashimoto I. and Ohno H. 2001. A new multivariate statistical process monitoring method using principal component analysis, Computer & Chemical Engineering, 25: 1103-1113.

19- Kisi O. 2009. Modeling monthly evaporation using two different neural computing techniques, Irrigation Science, 27:417–430.

20- 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.

21- Maheras P. 1984. Weather-type classification by factor analysis in Thessaloniki area, Climatol, 4:437-443.

22- Mohan S. and Arumugam N. 1996. Relative importance of meteorological variables in evapotranspiration: Factor analysis approach, Water Resource Management, 10: 1-20.

23- Morin G., Fortin J.P., Sochanska W.J., Lardeau P. and Charbonneau R. 1979. Use of principal component analysis to identify homogenous precipitation stations for optimal interpolation, Water Resource Reservoir, 15: 1841-1850.

24- Noori R., Khakpour A., Omidvar B. and Farokhnia A. 2010. Comparison of ANN and principal component analysis-multivariate linear regression models for predicting the river flow based on developed discrepancy ratio statistic, Expert Systems with Applications, 37: 5856-5862.

25- Ouyang Y. 2005. Application of principal component and factor analysis to evaluate surface water quality monitoring network, Water Reservoir, 39: 2621-2635.

26- 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 & Technology, 6: 467-476.

27- Qiu Y., Fu B., Wang J., Chen L. Meng Q. and Zhang Y. 2010. Spatial prediction of soil moisture content using multiple-linear regressions in a gully catchment of the Loess Plateau, China, Arid Environment, 74: 1-13.

28- Shirsath P.B. and Singh A.K. 2009. A comparative study of daily pan evaporation estimation using ANN, regression and climate based models, Water Resource Management, DOI 10.1007/s11269-009-9514-2.

29- Shrestha S. and Kazama F. 2007. Assesment of surface water quality using multivariate statistical techniques: A case study of the Fuji river basin, Japan, Environment Modelling Software, 22: 464- 475.

30- Simeonov V., Stratis J.A., Samara C., Zachariadis G., Voutsa D., Anthemidis A., Sofoniou M. and Kouimtzis Th. 2003. Assessment of the surface water quality in Northern Greece, Water Reservoir, 37: 4119-4124.

31- Storrie-Lombardi M.C., Irwin M.J., Hippel T. and Storrie-Lombardi L.J. 1994. Spectral classification with principal component analysis and artificial neural networks, Vistas in Astronomy, 38: 331-340.

32- Wang J., Barreto L., Wang L., Chen Y., Rishe N., Andrian J. and Adjouadi M. 2010. Multilinear principal component analysis for face recognition with fewer features, Neurocomp, 73: 1550-1555.

www.SID.ir

Archive of SID
Journal of Water and Soil
Vol. 24, No. 6, Jan-Feb 2011, p. 1186-1196

نشریه آب و خاک
جلد 24، شماره 6، بهمن - اسفند 1389، ص. 1196-1186

Developing a Combined Model of Multiple Linear Regression-Principal Componenet and Factor Analysis (MLR-PCA) for Estimation of Reference Evapotranspiration
(Case Study: Kerman Station)

A. Seifi[1]- S.M. Mirlatifi[2*]- H. Riahi[3]
Received: 11-4-2010
Accepted: 3-10-2010

Abstract

Reference evapotranspiration (ET_o) is an essential parameter required for proper management of agricultural crop irrigation. ET_o is influenced by many different hydrological variables and as a result is a very complex procces. ET_o is usually estimated by empirical or process-orinented models (mathematical relationships) from historical weather data. The need for accurate estimates of ET_o and the complexity of developing models to describe such complex process magnifies the need for developing new data mining methods. In this paper, the possibility of using a combined method of multiple linear regressions with principal componenets analysis (MLR-PCA) for estimating reference evapotranaspiration was investigated. In this analysis, measured daily meteorological data of Kerman synoptic weather station recorded from 1996 to 2005 were used. Three principal componenets that explained 80% of the total variance of the data were recognized as the principle componenets and others as disorder. Using the extracted principle componenets, a multiple linear regression model was developed to estimate ET_o. The statistic index of t for assessing the results of a fixed constant and each componenets of PC_1 and PC_2 were determined. According to the results, all coefficients were significant at the level of 95% and PC_1 had more importance than the other component namely PC2. This revealed that the variables of radiation intensity, relative humidity, sunshine hours, minimum temperature and maximum temperature had more importance in estimating reference evapotranspiration than other climatological parameters. Comparison of MLR-PCA model with Penman-Monteith results showed that about 82% of the total amount of the ET_o variance is defined by the three aformenstioned principle componenets.

Keywords: Reference Evapotranspiration, FAO Penman-Monteith, Multiple Regression, Principle Componenet Analysis.

1,2 -Graduate Student and Associate Professor, Department of irrigation and drainage engineering, Tarbiat Modares University
(*-Corresponding Author Email: mirlat_m@modares.ac.ir)
3-PhD Student, Department of Water Structures Engineering, Tarbiat Modares University

www.SID.ir