Introduction to multiple linear regression and its application to simulate daily pan evaporation

M. Jafari[1*], Y. Dinpashoh[2], E. Asadi[3]

1, 2, 3- MSc Student & Associate Professor & Assistant Professor, Department of Water Engineering, Faculty of Agriculture, University of Tabriz, Iran.

*(Corresponding Author Email: M.jafari.twone@gmail.com)

Received: 25-12-2015

Accepted: 7-2-2016

آشنایی با رگرسیون خطی چندگانه و استفاده از آن در شبیه‌سازی مقادیر تبخیر روزانه تشت

مجید جعفری[۱*]، یعقوب دین‌پژوه[۲]، اسماعیل اسدی[۳]

۱، ۲ و ۳- به ترتیب دانشجوی کارشناسی ارشد منابع آب، دانشیار و استادیار گروه مهندسی آب، دانشکده کشاورزی، دانشگاه تبریز.

*(نویسنده‌ی مسئول، E-Mail: M.jafari.twone@gmail.com)

تاریخ دریافت: ۹٤/۱۰/٤

تاریخ پذیرش: ۹٤/۱۱/۱۸

Abstract

Evaporation is one of the main parameters for the optimum operation of reservoirs, design of irrigation systems and scientific management of water resources. Accurate estimation of the water evaporation level is crucial in any region especially in arid and semiarid regions. In this study, the feasibility of simulation of pan evaporation in Maraghe station using the multiple regression models were investigated. Meteorological data, including the maximum and minimum air temperatures, dew point temperature, the maximum and minimum relative humidities, the number of sunshine hours, the wind speed records during the 1992 to 2012 (using the daily time scale) of Maraghe synoptic station were used. Various models of multiple linear regression (MLR) were derived for simulating the evaporation in the mentioned station. The selected MLR model was tested for multi-collinearity of input repressors using the Ridge Regression. For this purpose, the Variance Inflation Factor (VIF), responsible for the multi-collinearity in regression analysis, was calculated for each of the input variables. The results showed that all of the obtained VIFs values were less than 10 and the multi-collinearity is not created. Furthermore, the ratio of eigenvalues of the correlation matrix, $\lambda_{max}\backslash\lambda_{min}$, for the selected model was calculated as 6.2. The selected model consisted of the maximum relative humidity (RH_{max}), number of sunshine hours (n), and minimum air temperature (T_{min}) as independent inputs, and the Pan evaporation as the dependent variable, f (RH_{max}, T_{min}, n). Therefore, it can be concluded that there was no multi-collinearity in the selected MLR model. The RMSE and R^2 values of the selected model (MLR) was calculated as 2.37 and 0.676 mm/day, respectively.

Keywords: Evaporation, Regression Models, Multi-collinearity, Ridge Regression, Synoptic station of Maraghe.

چکیده

تبخیر یکی از پارامترهای اصلی در برنامه‌ریزی و بهره‌برداری بهینه از مخازن، طراحی سیستم‌های آبیاری و مدیریت علمی منابع آب می‌باشد. تخمین دقیق تبخیر، در هر منطقه به ویژه در نواحی خشک و نیمه خشک از اهمیت بالایی دارد. در این مطالعه، امکان شبیه‌سازی مقدار تبخیر از تشت با استفاده از مدل‌های رگرسیون چندگانه بررسی شد. داده‌های هواشناسی شامل حداکثر دمای هوا، حداقل دمای هوا، دمای نقطه شبنم، حداکثر رطوبت نسبی هوا، حداقل رطوبت نسبی هوا، تعداد ساعات آفتابی و سرعت باد در مقیاس روزانه در دوره آماری ۱۳۹۱-۱۳۷۱ در ایستگاه سینوپتیک مراغه استفاده شد. مدل رگرسیونی چندگانه خطی (MLR)[۱] برای ایستگاه مذکور بسط داده شد. مدل منتخب MLR توسط روش رگرسیون ریچ به منظور لحاظ کردن اثر همخطی توام بین ورودی‌های مدل مورد آزمون واقع شد. مقادیر ضریب تورم واریانس (VIF)[۲] که شاخصی برای بررسی همخطی توام در تحلیل رگرسیون است، برای یکایک متغیرها محاسبه شد. نتایج نشان داد که مقدار عددی همه ضرایب مذکور کمتر از ۱۰ بوده و هیچ متغیری ایجاد همخطی نکرد. افزون بر این، مقدار نسبت مقادیر ویژه ماتریس همبستگی ($\lambda_{max}\backslash\lambda_{min}$)، برای مدل منتخب سه متغیره f(RH_{max}, T_{min}, n)، تابعی است سه متغیره که متغیرهای مستقل آن حداکثر رطوبت نسبی (RH_{max})، تعداد ساعات آفتابی (n) و حداقل دمای هوا (T_{min}) می‌باشد برابر با ۶/۲ بدست آمد. بنابراین، همخطی توام در مدل MLR منتخب مشاهده نشد. مقادیر RMSE و R^2 برای مدل منتخب MLR به ترتیب برابر ۲/۳۷ میلی‌متر بر روز و ۰/۶۷۶ بدست آمد.

واژه‌های کلیدی: تبخیر، مدل‌های رگرسیونی، هم‌خطی توام، رگرسیون ریچ، ایستگاه سینوپتیک مراغه.

مقدمه

تبخیر یکی از مولفه‌های اصلی چرخه‌ی هیدرولوژی و توازن آب در طبیعت و زیست‌بوم‌های کشاورزی است. در پدیده‌ی تبخیر عوامل متعددی دخالت دارند که از مهم‌ترین آنها می‌توان به دمای هوا، سرعت باد، رطوبت نسبی، تابش خورشیدی و ارتفاع محل اشاره کرد (Chow و همکاران، ۱۹۸۸). کاربرد مفاهیم آماری به عنوان یک اصل زیربنایی در تحلیل بسیاری از مسایل پیچیده تصمیم‌گیری کاملاً پذیرفته شده است. هدف اینگونه تحلیل‌ها در شکل کلی آن پیدا کردن ارتباط بین متغیر وابسته و متغیرهای مستقل است. با اینکه ضریب همبستگی[۳] نشان‌دهنده ارتباط بین دو متغیر است اما زمانی که هدف، اثر یك متغیر بر متغیر دیگر است، نمی‌توان از آن استفاده کرد. از طرفی گاهی تخمین تغییر در یك متغیر با تغییر متغیر دیگر مهم است که باز هم ضریب همبستگی نمی‌تواند تخمینی از این تغییرات ارائه دهد. برای پاسخ به سئوالات فوق باید به سراغ رگرسیون رفت.

یکی از روش‌های آماری، روش رگرسیون چند متغیره خطی است. در رگرسیون یك متغیر که تأثیرپذیر (Y) از سایر متغیرهاست، متغیر پاسخ/متغیر وابسته نامیده می‌شود. متغیر یا متغیرهایی که بر متغیر پاسخ اثر می‌گذارند (X) متغیر توضیحی/متغیر مستقل نامیده می‌شود. Shirsath و Kumar (۲۰۰۹) تبخیر روزانه از تشت را با روش‌های شبکه‌های عصبی مصنوعی (ANN)[۴]، MLR و مدل‌های پنمن، استفن[۵] و استووارت[۶] پیش‌بینی کردند. نتایج نشان داد که روش‌های ANN و MLR در مقایسه با سایر روش‌ها کارایی بهتری داشتند. مطالعاتی توسط Bruton و همکاران (۲۰۰۰)، Shirsath و Singh (۲۰۱۰)، Shirgure (۲۰۱۱)، Googhari (۲۰۱۲) و Shirgure و Rajput (۲۰۱۲) با استفاده از روش‌های ANN و MLR انجام گرفت که نشان دهنده نتایج قابل قبول برای هر دو مدل و البته نتایج قابل اعتمادتر برای ANN می‌باشد. Almedeij (۲۰۱۲) پیش‌بینی تبخیر روزانه و ماهانه را در کویت توسط مدل رگرسیون خطی انجام داد. برای این کار پارامترهای دما، رطوبت نسبی هوا و سرعت باد در دوره زمانی ۱۷ ساله (از ژانویه ۱۹۹۳ تا دسامبر ۲۰۰۹) در یک منطقه بیابانی بکار برده شدند. در نهایت رابطه بین تبخیر با دما و رطوبت نسبی هوا بصورت خطی ایجاد شد. نتایج وی نشان داد که مدل ایجاد شده با این پارامترها همبستگی بسیار بالایی با داده‌های مشاهداتی دارد. Ladlani و همکاران (۲۰۱۲) تبخیر-تعرق روزانه در منطقه مدیترانه‌ای کشور الجزایر را توسط روش‌های نروفازی و MLR مدل‌سازی کردند. داده‌های بکار رفته شامل میانگین رطوبت نسبی روزانه، ساعات آفتابی، حداکثر، حداقل و میانگین دمای هوا و سرعت باد بودند. نتایج بدست آمده نشان داد که عملکرد هر دو مدل در پیش‌بینی تبخیر-تعرق قابل قبول است. Mallik و همکاران (۲۰۱۳) و Kisi (۲۰۰۹)، برای برآورد تبخیر از تشت از روش‌های ANN و MLR استفاده کردند. نتایج این دو مطالعه نشان دهنده کارایی خوب مدل MLR بوده است. Malik و Kumar (۲۰۱۵) تبخیر روزانه از تشت را توسط روش‌های ANN، نروفازی و MLR در منطقه پانتاگار (هند) شبیه‌سازی کردند. مقایسه نتایج با استفاده از معیارهای ارزیابی مانند ضریب تعیین (R^2) و مجذور میانگین مربعات خطا (RMSE) نشان دهنده کارایی بهتر روش‌ها به ترتیب ANN، نروفازی و MLR بود. اسکافی نوغانی و همکاران (۱۳۸۷) از روش رگرسیون خطی چندگانه (MLR) با استفاده از پارامترهای هواشناسی، مقدار تبخیر از تشت را برآورد نمودند و با مقادیر اندازه‌گیری شده از تشت تبخیر مقایسه کردند. منطقه مورد مطالعه ایشان حوضه گرگان‌رود بود. نتایج نشان داد که خروجی روش MLR در این حوضه با مقادیر اندازه‌گیری شده تقریباً برابر بوده و بنابراین مدل دقت مناسب در تخمین تبخیر را دارد.

مدل‌های رگرسیونی در حوزه‌های مختلف علمی کاربرد وسیعی دارند. مسئله جدی که می‌تواند استفاده از مدل رگرسیونی را با محدودیت مواجه کند، وجود هم‌خطی توام (یا چندگانه) بین متغیرهای رگرسیونی است. وجود هم‌خطی توام، توانایی برآورد دقیق ضرایب مدل رگرسیون را با مشکل مواجه می‌کند. در شرایط وجود هم‌خطی توام تفسیر ضرایب رگرسیون قرین صحت نخواهد بود. بنابراین برای از بین بردن هم‌خطی، از روش رگرسیون ریج استفاده شد. Li و همکاران (۲۰۱۰) سیستمی با دو سری داده متفاوت برای مقایسه دو روش رگرسیون خطی ساده و رگرسیون خطی ریج ارائه دادند. نتایج نشان‌دهنده عملکرد بهتر رگرسیون ریج در رگرسیون‌بندی داده‌های دارای هم‌خطی توام و نتایج قابل شرح‌تر نسبت به رگرسیون خطی ساده است. فکری و همکاران (۱۳۹۰) رگرسیون ریج را به عنوان یکی از کاراترین روش‌های مقابله با مشکل هم‌خطی در داده‌های مربوط به عوامل خطر بیماری‌های قلبی معرفی کردند.

با توجه به اهمیت شبیه‌سازی پدیده‌های اقلیمی، هدف از مقاله حاضر آشنایی با روش‌های مختلف رگرسیون چندگانه و تعیین کارایی آنها در تخمین مقادیر تبخیر از تشت روزانه می‌باشد. بنابراین در تحقیق حاضر به عنوان مطالعه موردی، سری زمانی تبخیر روزانه ایستگاه مراغه با روش‌های رگرسیون چندگانه خطی و رگرسیون ریج مدل‌بندی شدند.

مواد و روش‌ها

• منطقه مورد مطالعه

شهرستان مراغه با وسعت ۲۵۹۷ کیلومتر مربع (۴/۸٪ مساحت استان آذربایجان شرقی) در ۱۳۰ کیلومتری تبریز واقع شده و در عرض شمالی ۳۷ درجه و ۱ دقیقه الی ۳۷ درجه و ۴۵ دقیقه و در طول شرقی ۴۶ درجه و ۹ دقیقه الی ۴۶ درجه و ۴۴ دقیقه قرار گرفته است. بخش شمالی که شامل دامنه‌های جنوبی ارتفاعات سهند می‌باشد، به صورت کوهستانی و ناهموار بوده و بخش مرکزی و جنوبی شهرستان، به صورت دشت و جلگه می‌باشد. میزان بارش باران و تبخیر در این استان به ترتیب بطور متوسط حدود ۳۵۰ و ۱۷۰۰ میلی‌متر در سال است. فصول زمستان و بهار، فصول بارندگی منطقه محسوب می‌شوند و بیشترین شدت بارندگی در بهار مشاهده می‌شود.

• داده‌های مورد استفاده

در این مطالعه از داده‌های روزانه تبخیر از تشت (E)، دمای حداکثر (T_{max}) و دمای حداقل (T_{min})، رطوبت نسبی حداکثر (RH_{max}) و رطوبت نسبی حداقل (RH_{min})، ساعات آفتابی (n)، سرعت باد (wind) استفاده شده است. داده‌ها از سازمان آب منطقه‌ای آذربایجان شرقی اخذ شده و پس از بررسی اشتراک زمانی داده‌های ثبت شده برای پارامترهای ذکر شده در بالا، سال ۱۳۷۱ به عنوان مبدأ مطالعات انتخاب شد. بدلیل فقدان یا گم‌شدن برخی داده‌ها، از روش رگرسیون خطی ساده ما بین سری زمانی داده‌های مراغه و ایستگاه‌های همسایه برای تخمین این داده‌ها استفاده شد. بعد از تشکیل ماتریس همبستگی، معنی‌داری ضرایب همبستگی بین ایستگاه‌های مختلف توسط آزمون t در سطح معنی‌داری ۵٪ مورد آزمون واقع شد. سپس مناسب‌ترین همبستگی بین داده‌های ایستگاه مراغه و سایر ایستگاه‌ها در نظر گرفته شد و داده‌های گمشده تخمین زده شدند. آزمون t برای بررسی معنی‌داری همبستگی بین متغیرها، به شرح رابطه زیر است (Hirsch و همکاران، ۱۹۹۳):

$$t = \frac{r.\sqrt{n-2}}{\sqrt{1-r^2}} \qquad (۱)$$

که در آن r ضریب همبستگی بین دو متغیر و n تعداد داده‌ها می‌باشد. اگر قدر مطلق t محاسبه شده بیشتر از مقدار t جدول (جدول t-student با درجه آزادی n-2 و سطح معنی داری $\alpha/2$) باشد، آنگاه ضریب همبستگی معنی‌دار تلقی می‌گردد. بعد از جاگذاری داده‌های گم‌شده (برای پارامترهای دما و رطوبت نسبی هوا)، کیفیت داده‌ها کنترل شد. برای این منظور در طول کل دوره آماری، T_{max} و T_{min} در یک نمودار و همینطور RH_{max} و RH_{min} در نموداری دیگر رسم شد. در هر نمودار بیشتر بودن سری نظیر حداکثر نسبت به سری حداقل کنترل می‌شد، به عبارت دیگر از بیشتر بودن مقدار پارامتر حداکثر نسبت به مقدار حداقل در همان روز اطمینان حاصل می‌شد. پس از کنترل کیفی داده‌ها مقادیر تبخیر از تشت به عنوان متغیر وابسته و مقادیر متغیرهای T_{max}، T_{min}، RH_{max}، RH_{min}، n و wind به عنوان متغیرهای مستقل در نظر گرفته شدند. برای اینکه داده‌ها دارای دامنه تعریف متفاوتی هستند داده‌های هر سری با رابطه زیر استاندارد شدند.

$$Z = \frac{x-\mu}{\sigma} \qquad (۲)$$

که در آن، Z متغیر استاندارد شده، X مقدار متغیر مشاهداتی، μ میانگین و σ نیز انحراف معیار داده‌های آن متغیر است.

• مدل رگرسیون خطی چندگانه (MLR)

ارتباط خطی بین متغیرهای مستقل $x_1, x_2, \dots, x_n$ و متغیر Y در مدل MLR، به شرح زیر است:

$$Y=a_0 +a_1x_1+ a_2x_2+ \dots + a_nx_n +e \qquad (۳)$$

که در آن، a_0 عرض از مبدأ و پارامترهای $a_1, a_2, \dots, a_n$ ضرایب رگرسیونی می‌باشند (Balan و همکاران، ۱۹۹۵).

همچنین مدل MLR در فرم ماتریسی به شرح زیر نوشته می‌شود:

$$Y = X\beta + \varepsilon \qquad (۴)$$

که در آن β: بردار ضرایب رگرسیون (P+1×1)، ε: ماتریس خطای برازش (n×1)، Y: بردار نظیر متغیر وابسته (n×1)، X: ماتریس متغیرهای مستقل (n×P+1)، n تعداد داده‌ها و p تعداد متغیرها می‌باشند که به شرح زیر تعریف می‌شوند:

$$Y=\begin{bmatrix} Y_1 \\ Y_2 \\ . \\ . \\ . \\ Y_n \end{bmatrix}, X=\begin{bmatrix} 1 & x_{11} & x_{12} & . & . & . & x_{1p} \\ 1 & x_{21} & x_{22} & . & . & . & x_{2p} \\ . & . & . & . & . & . & . \\ . & . & . & . & . & . & . \\ . & . & . & . & . & . & . \\ 1 & x_{n1} & x_{n2} & . & . & . & x_{np} \end{bmatrix}, \beta=\begin{bmatrix} \beta_0 \\ \beta_1 \\ . \\ . \\ . \\ \beta_p \end{bmatrix}, \varepsilon=\begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ . \\ . \\ . \\ \varepsilon_n \end{bmatrix} \qquad (۵)$$

با حل رابطه بالا برحسب β خواهیم داشت:

$$\hat{\beta} = (X'X)^{-1} (X'Y) \qquad (۶)$$

که در آن X' ترانهاده ماتریس X است. برای محاسبه معکوس ماتریس ($X'X$)، لازم است دترمینان عبارت فوق صفر نگردد (خداشناس و همکاران ۱۳۸۷، نوری و همکاران ۱۳۸۷). در این مطالعه مدل MLR هم برای متغیرهای مشاهداتی و هم برای متغیرهای استاندارد شده بسط داده

شد. افزون بر این جهت لحاظ کردن اثر وابستگی متغیرهای مستقل با همدیگر (هم‌خطی توام)، مدل رگرسیون ریج[۶] (تیغه‌ای) نیز بکار برده شد.

• هم‌خطی توام

مسئله جدی که می‌تواند استفاده از مدل رگرسیونی را با اشکال مواجه کند، هم‌خطی چندگانه یا وابستگی خطی نزدیک بین متغیرهای رگرسیونی است. وجود وابستگی خطی نزدیک، توانایی برآورد دقیق و صحیح ضرایب مدل رگرسیون را با مشکل مواجه می‌کند. اگر بین متغیرهای مستقل وابستگی (همبستگی معنی‌دار) وجود داشته باشد، در این حالت استفاده از MLR بدون درنظر گرفتن اثر مذکور (هم‌خطی توام) می‌تواند گمراه کننده باشد. برای پی بردن به وجود حالت هم‌خطی توام چند راه وجود دارد. روش بکار برده شده در این مقاله بشرح زیر است (Bowker و Lieberman، ۱۹۷۲).

۱- استفاده از ضرایب تورم واریانس (VIF)، که به شرح زیر محاسبه می‌شوند:

$$VIF(b_j)=\frac{1}{(1-R_j^{\,2})},\quad j=1,2,\ldots p \qquad (۷)$$

که در آن R^2_j ضریب تعیین ناشی از رگرسیون چندگانه خطی با در نظر گرفتن x_j به عنوان متغیر وابسته و p-1 متغیر دیگر به عنوان متغیرهای مستقل می‌باشند. b_jها ضرایب اصلی رگرسیون در این رابطه هستند. به تعداد متغیرها (P) ضرایب (VIF) محاسبه می‌شوند و در حقیقت ضرایب تورم واریانس‌ها عناصر قطر اصلی وارون ماتریس همبستگی هستند. در این مطالعه هر متغیری که ضریب VIF آن بیشتر از ۱۰ بدست آید، به عنوان متغیری که باعث هم‌خطی توام می‌شود در نظر گرفته می‌شود (Bowker و Lieberman، ۱۹۷۲).

۲- مقادیر ویژه یا (ریشه‌های مشخصه) ماتریس همبستگی، معیاری مناسب برای سنجش هم‌خطی توام هستند. اگر $X'X$ در شکل همبستگی باشد، مقادیر ویژه ماتریس $X'X$ در واقع ریشه‌های معادله رابطه (۸) هستند.

$$|X'X\text{-}\lambda I|=0 \qquad (۸)$$

وجود یک یا بیشتر از یک مقدار ویژه نزدیک به صفر بدین معناست که شرط هم‌خطی توام موجود است. اگر λ_{max} و λ_{min} معرف بزرگترین و کوچکترین مقادیر ویژه $X'X$ باشند، نسبت $\lambda_{max}\backslash\lambda_{min}$ به عنوان معیاری مناسب برای سنجش هم‌خطی توام است. هرچه مقدار این نسبت بیشتر باشد، درجه هم‌خطی بیشتر است. بطور کلی اگر این نسبت از ۱۰ کمتر باشد هم‌خطی توام قابل صرف نظر کردن است (Bowker و Lieberman، ۱۹۷۲).

• رگرسیون ریج

برای حل مساله هم‌خطی توام چند روش مختلف پیشنهاد شده است. در این مطالعه به جای استفاده از روش حداقل مربعات، برای تخمین ضرایب مدل رگرسیون خطی چندگانه از روش رگرسیون ریج استفاده شد. در این مدل برای برآورد ضرایب مدل رگرسیون ریج رابطه زیر در نظر گرفته شد (Li و همکاران، ۲۰۱۰):

$$B^*(k)=(X'X+kI)^{-1}X'Y \qquad (۹)$$

که در آن I ماتریس یکه به ابعاد p×p، $\mathbf{B}^*(k)$ برآوردکننده تیغه‌ای یا همان ضرایب رگرسیون ریج و kها مقادیر بهینه‌گر ضرایب رگرسیون ریج هستند تا هم‌خطی توام را در صورت وجود از بین ببرند. به‌طورکلی برای هر مسأله یک مقدار k بهینه وجود دارد، که ساده‌ترین روش یافتن آن حل معادله (۹) به ازای چند مقدار k واقع در فاصله $0 \le k \le 1$ است. در این مطالعه مقادیر ۰/۰۵، ۰/۱، ...، ۱ برای k در نظر گرفته شد و سپس نمودار مقادیر $\mathbf{B}^*(k)$ تابعی از k در یک دیاگرام رسم شد. این نمودار، مسیر تیغه نامیده می‌شود. مقدار k چنان برگزیده شد که برآوردهای نسبتاً پایدار برای پارامترها ($\mathbf{B}^*(k)$) بدست آید. بعبارت بهتر مقادیر ضرایب با افزایش k تغییر چندانی نداشته باشد (نمودار تقریباً افقی باشد). نحوه استانداردسازی داده‌ها در روش رگرسیونی ریج نیز به شرح زیر در نظر گرفته شد:

$$Z_{ij}=\frac{X_{ij}-\bar{X}_j}{\sqrt{S_{jj}}}\quad i=1,2,\ldots,n\,،\,j=1,2,\ldots,p \qquad (۱۰)$$

$$S_{jj}=\sum_{i=1}^{n}(X_{ij}-\bar{X}_j)^2 \qquad (۱۱)$$

• معیارهای ارزیابی عملکرد مدل‌ها

به منظور مقایسه و ارزیابی عملکرد مدل‌های مورد بررسی از پارامترهای میانگین مربعات خطا (RMSE)، خطای مطلق میانگین (MAE)، ضریب تعیین (R^2) استفاده گردید. فرمول‌های استفاده شده برای این پارامترها به صورت زیر بود:

$$RMSE=(\frac{1}{n}\sum_{i=1}^{n}(p_i-o_i)^2)^{0.5} \qquad (۱۲)$$

$$MAE=\frac{1}{n}\sum_{i=1}^{n}|p_i-o_i| \qquad (۱۳)$$

$$R^2=\frac{(\sum_{i=1}^{n}(\mathrm{Pi}-\bar{P})(\mathrm{Oi}-\bar{O}))^2}{\sum_{i=1}^{n}(\mathrm{Pi}-\bar{P})^2\sum_{i=1}^{n}(\mathrm{Oi}-\bar{O})^2} \qquad (۱۴)$$

در این روابط، n تعداد داده‌ها، O_i مقادیر مشاهداتی تبخیر از تشت، P_i مقادیر محاسباتی تبخیر از مدل‌ها می‌باشد. بعد RMSE و MAE میلی‌متر بر روز و R^2 بدون بعد است.

هرچه R^2 به عدد یک نزدیک شود عملکرد مدل خوب ارزیابی می‌گردد. در این حالت معمولاً مقادیر آماره‌های RMSE و MAE به سمت صفر میل می‌کنند (Salas، ۱۹۹۳).

نتایج و بحث

در این مطالعه ابتدا تعداد ۵ ماه (از اول آبان تا آخر اسفند) از سال به علت یخبندان یا عدم ثبت داده نظیر تبخیر در تمام سال‌های مربوط به دوره آماری حذف گردید و سایر داده‌های تبخیر از تشت بصورت روزانه (از اول فروردین سال ۱۳۷۱ تا آخر مهر همان سال) در صفحه اکسل درج شد. سپس مدل MLR بین E و متغیرهای هواشناسی (با اطلاعات آماری مندرج در جدول ۱) در دو حالت انجام شد. در حالت اول متغیرهای مستقل و وابسته بدون استاندارد شدن وارد مدل گردیدند و ضرایب پارامترهای مدل و ضریب تعیین بدست آمد. در حالت دوم متغیرهای مذکور توسط رابطه (۱۰) استاندارد و سپس وارد مدل شدند و ضرایب پارامترهای مدل و ضریب تعیین بدست آمد. آنگاه بهترین مدل از روی ارزیابی عملکرد آن انتخاب شد.

جدول (۲) مقادیر ماتریس همبستگی بین همه متغیرهای ورودی به مدل را نشان می‌دهد. این مقادیر میزان همبستگی متغیرها را نسبت بهم نشان می‌دهد.. معنی‌داری همبستگی‌ها با آزمون t استیودنت بررسی شده، برای همه آنها در سطح ۵٪ اثبات شد.

جدول ۱- اطلاعات نمودار جعبه‌ای متغیرهای بکار رفته برای ایستگاه مراغه در مطالعه حاضر (۱۳۹۱- ۱۳۷۱)

متغیرها	ماکزیمم	مینیمم	چارک اول	چارک سوم	میانه	میانگین	انحراف معیار
تبخیر (mm/day)	۲۱/۳	۱	۵/۳	۱۲/۱	۸/۸	۸/۷۷	۴/۱۷
دمای حداقل هوا (C°)	۲۸/۴	-۴/۶	۹/۱۵	۱۸/۴	۱۴	۱۳/۵۶	۵/۹۳
دمای حداکثر هوا (C°)	۴۰/۴	۴/۲	۲۰/۶	۳۱/۸	۲۷	۲۵/۹۲	۷/۰۹
رطوبت نسبی حداکثر	۱۰۰	۱۷	۴۲	۶۷	۵۴	۵۵/۶۹	۱۷/۴۶
رطوبت نسبی حداقل	۵۲	۱	۱۵	۳۰	۲۱	۲۴/۰۸	۱۲/۱۲
سرعت باد (m/s)	۸/۴	۰	۱/۹	۴/۵	۳/۱۰	۳/۲۸	۱/۸۵
ساعات آفتابی (hr)	۱۴/۲	۳/۷	۸/۸	۱۲/۲	۱۰/۹	۹/۸۹	۳/۲۴

جدول ۲- ماتریس متقارن ضرایب همبستگی بین متغیرهای هواشناسی در ایستگاه مراغه

نام متغیر	دمای حداکثر	دمای حداقل	حداکثر رطوبت نسبی	حداقل رطوبت نسبی	تعداد ساعات آفتابی	سرعت باد
دمای حداکثر	۱	۰/۹۳۱	-۰/۷۴۶	-۰/۶۷۸	۰/۶۲۳	۰/۲۷۸
دمای حداقل	۰/۹۳۱	۱	-۰/۶۴۴	-۰/۴۹۶	۰/۴۵۲	۰/۳۸۶
حداکثر رطوبت نسبی	-۰/۷۴۶	-۰/۶۴۴	۱	۰/۷۷۶	-۰/۵۴۵	-۰/۱۱۴
حداقل رطوبت نسبی	-۰/۶۷۸	-۰/۴۹۶	۰/۷۷۶	۱	-۰/۶۳۳	۰/۰۳۵
تعداد ساعات آفتابی	۰/۶۲۳	۰/۴۵۲	-۰/۵۴۵	-۰/۶۳۳	۱	۰/۱
سرعت باد	۰/۲۷۸	۰/۳۸۶	-۰/۱۱۴	۰/۰۳۵	۰/۱	۱

جدول (۳) و (۴) نتایج مدل‌های مختلف رگرسیون چندگانه خطی را نشان می‌دهد که به ترتیب برای داده‌های مشاهده‌ای و داده‌های استاندارد شده با رابطه (۲) ایجاد شده‌اند. از این جداول می‌توان استنباط کرد، مدل‌ها به ترتیب به صورت پسرو و با درجه اهمیت نزولی ردیف شده‌اند. در جدول (۳) متغیرهای مورد نیاز برای مدل ۱ شامل دمای حداقل هوا، رطوبت حداکثر و حداقل هوا، سرعت باد و ساعات آفتابی می‌باشد که عرض از مبدأ و ضرایب b_1 تا b_6 هر مدل به همراه آماره‌های عملکرد نشان داده شده‌اند. مقادیر آماره R^2 در مدل‌های مختلف بین ۰/۶۲۶ در مدل ۷ تا ۰/۶۸۲ در مدل ۱ تغییر می‌کند. مدل‌هایی که دارای R^2 بالایی هستند RMSE و MAE پایینی را نشان می‌دهند. مقادیر RMSE از ۲/۳۵ در مدل ۱ تا ۲/۵۵ میلی‌متر بر روز در مدل ۷ تغییر می‌کند.

ملاحظه می‌شود که با کاهش تعداد متغیرهای ورودی برای مدل‌ها در روش رگرسیون خطی میزان خطا افزایش و میزان R^2 کاهش می‌یابد، ولی این افزایش خطا یا کاهش ضریب تعیین با توجه به ارقام موجود در ستون‌های RMSE، MAE و R^2 در هرکدام از جداول (۳) و (۴) ناچیز و در حدود ۵ الی ۱۰ درصد بین اولین مدل با بیشترین متغیر و آخرین مدل با فقط یک متغیر می‌باشد. لذا با بررسی اعداد و ارقام موجود در این جداول مدل شماره ۳ با (R^2=۰/۶۷۶) با ۳ متغیر مورد نیاز برای محاسبه تبخیر از تشت (هم در حالتی که داده‌ها استاندارد شده و هم برای داده‌های معمولی) مناسب می‌باشد. متغیرهای مورد نیاز برای مدل رگرسیون خطی شماره ۳ عبارتند از حداکثر رطوبت نسبی هوا، حداقل دمای هوا و ساعات آفتابی می‌باشند.

جدول ۳- ضرایب و آماره‌های عملکرد مدل‌های رگرسیونی چندگانه خطی مختلف با استفاده از داده‌های غیر استاندارد

مدل‌ها	b_0	b_1	b_2	b_3	b_4	b_5	b_6	R^2	RMSE (mm/day)	MAE
مدل ۱	۳/۰۴۶	۰/۰۰۰	۰/۴۰۰	-۰/۰۳۶	-۰/۰۱۳	۰/۱۹۲	۰/۲۰۳	۰/۶۸۲	۲/۳۵۰	۱/۸۰
مدل ۲	۲/۹۳۰	۰/۰۰۰	۰/۴۰۱	-۰/۰۴۲	۰/۰۰۰	۰/۱۸	۰/۲۱۷	۰/۶۸۱	۲/۳۵۵	۱/۸۰
مدل ۳ (منتخب)	۲/۸۹۹	۰/۰۰۰	۰/۴۳۱	-۰/۰۳۸	۰/۰۰۰	۰/۰۰۰	۰/۲۱۵	۰/۶۷۶	۲/۳۷۳	۱/۸۱
مدل ۴	۳/۰۸۱	-۰/۰۱۴	۰/۴۴۳	-۰/۰۳۹	۰/۰۰۰	۰/۰۰۰	۰/۲۲۰	۰/۶۷۶	۲/۳۷۳	۱/۸۱
مدل ۵	-۱/۷۱۰	۰/۰۸۰	۰/۴۰۸	۰/۰۰۰	۰/۰۰۰	۰/۰۰۰	۰/۲۳۵	۰/۶۶۵	۲/۴۱۰	۱/۸۵
مدل ۶	-۱/۳۹۵	۰/۲۴۳	۰/۲۸۶	۰/۰۰۰	۰/۰۰۰	۰/۰۰۰	۰/۰۰۰	۰/۶۴۸	۲/۴۷۰	۱/۹۰
مدل ۷	-۳/۲۸۷	۰/۴۶۵	۰/۰۰۰	۰/۰۰۰	۰/۰۰۰	۰/۰۰۰	۰/۰۰۰	۰/۶۲۶	۲/۵۵۰	۱/۹۷

توجه: b_0 عرض از مبدا و b_1 تا b_6 به ترتیب ضرایب پارامترهای (Tmax، Tmin، RHmax، RHmin، wind، n) می‌باشد.

جدول ۴- ضرایب و آماره‌های عملکرد مدل‌های رگرسیونی چندگانه خطی مختلف با استفاده از داده‌های استاندارد

مدل‌ها	b_0	b_1	b_2	b_3	b_4	b_5	b_6	R^2	RMSE (mm/day)	MAE
مدل ۱	۰/۰۰۰	۰/۰۰۰	۰/۵۷۰	-۰/۱۵۲	-۰/۰۳۸	۰/۰۸۵	۰/۱۵۸	۰/۶۸۲	۰/۵۶۴	۰/۴۳
مدل۲	۰/۰۰۰	۰/۰۰۰	۰/۵۷۰	-۰/۱۷۶	۰۰۰/۰	۰/۰۸۰	۰/۱۶۹	۰/۶۸۱	۰/۵۶۵	۰/۴۳
مدل۳	۰/۰۰۰	۰/۰۰۰	۰/۶۱۳	-۰/۱۵۸	۰۰۰/۰	۰/۰۰۰	۰/۱۶۸	۰/۶۷۶	۰/۵۶۹	۰/۴۴
مدل۴	۰/۰۰۰	-۰/۰۲۳	۰/۶۳۱	-۰/۱۶۲	۰/۰۰۰	۰/۰۰۰	۰/۱۷۲	۰/۶۷۶	۰/۵۶۹	۰/۴۴
مدل۵	۰/۰۰	۰/۱۳۶	۰/۵۸۱	۰/۰۰۰	۰/۰۰۰	۰/۰۰۰	۰/۱۸۳	۰/۶۶۵	۰/۵۷۹	۰/۴۴
مدل ۶	۰/۰۰۰	۰/۴۱۳	۰/۴۰۷	۰/۰۰۰	۰/۰۰۰	۰/۰۰۰	۰/۰۰۰	۰/۶۴۸	۰/۵۹۳	۰/۴۶
مدل ۷	۰/۰۰۰	۰/۷۹۱	۰/۰	۰/۰۰۰	۰/۰۰۰	۰/۰۰۰	۰/۰۰۰	۰/۶۲۶	۰/۶۱۱	۰/۴۷

توجه: b_0 عرض از مبدا و b_1 تا b_6 به ترتیب ضرایب پارامترهای (Tmax، Tmin، RHmax، RHmin، wind، n) می‌باشد.

برای پی بردن به وجود حالت هم‌خطی توام از مقادیر ویژه ماتریس همبستگی (λ) و عناصر روی قطر اصلی وارون ماتریس همبستگی $(X'X)^{-1}$ در مدل منتخب استفاده شد. نتایج نشان داد که ماتریس $(X'X)$ و معکوس آن یعنی $(X'X)^{-1}$ در مدل منتخب رگرسیون خطی چندگانه $f\ (RH_{max}, T_{min}, n)$ به شرح رابطه (۱۵) می‌باشد:

$$\mathbf{X'X} = \begin{bmatrix} 1 & 0.451825 & -0.6441 \\ 0.451825 & 1 & -0.54532 \\ -0.6441 & -0.54532 & 1 \end{bmatrix} \qquad (۱۵)$$

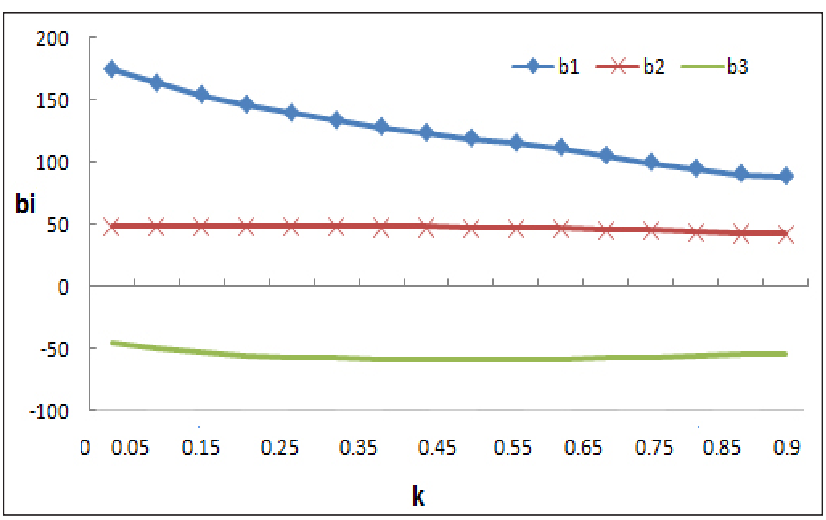
$$(\mathbf{X'X})^{-1} = \begin{bmatrix} 1.752107 & -0.25082 & 0.991745 \\ -0.25082 & 1.459146 & 0.634152 \\ 0.991745 & 0.634152 & 1.984596 \end{bmatrix}$$

مقادیر ویژه ماتریس $(X'X)$ با استفاده از نرم‌افزار Maple به شرح $\lambda_1 = ۲/۰۹۸$ و $\lambda_2 = ۰/۳۴$ و $\lambda_3 = ۰/۵۶۲$ بدست آمد؛ بنابراین، نسبت $\lambda_{max} \backslash \lambda_{min}$ برابر با ۶/۲ حاصل شد. باتوجه به اینکه نسبت مذکور کمتر از ۱۰ می‌باشد، می‌توان نتیجه گرفت که احتمال وجود هم‌خطی توام در پارامترهای مربوط به مدل رگرسیونی منتخب ناچیز می‌باشد. با این حال مدل رگرسیونی ریچ نیز به منظور بررسی نهایی محاسبه و انطباق داده می‌شود و فقط مدل منتخب آورده شده است. در جدول (۵) مقادیر ضرایب رگرسیون تیغه‌ای به ازای مقادیر مختلف k که با استفاده از رابطه‌ی (۹) محاسبه شده است، ارائه شده است. شکل (۲) نیز مسیر تیغه‌ای ضرایب رگرسیونی b_1، b_2 و b_3 را به ازای مقادیر مختلف k نشان می‌دهد. بطوری که از شکل (۳) می‌توان استنباط کرد به ازای مقدار k معادل با ۰/۰۵ مسیر تیغه‌ها نسبتاً بهم نزدیک و افقی شده است. از آنجاییکه طبق نظریه Bowker و Lieberman (۱۹۷۲)، در نقطه‌ای که نمودار شروع به افقی شدن می‌کند برآوردهای نسبتاً پایدار و نتایج قابل اعتمادتری حاصل خواهد شد، با استفاده از رقم مذکور و نیز $b_0 = ۸/۷۷۳$ مدل رگرسیون تیغه‌ای به شرح زیر حاصل شد:

$$E = 8.77 + 164.1076\, T_{min} + 48.03529 n - 49.8627\, RH_{max} \qquad (۱۵)$$

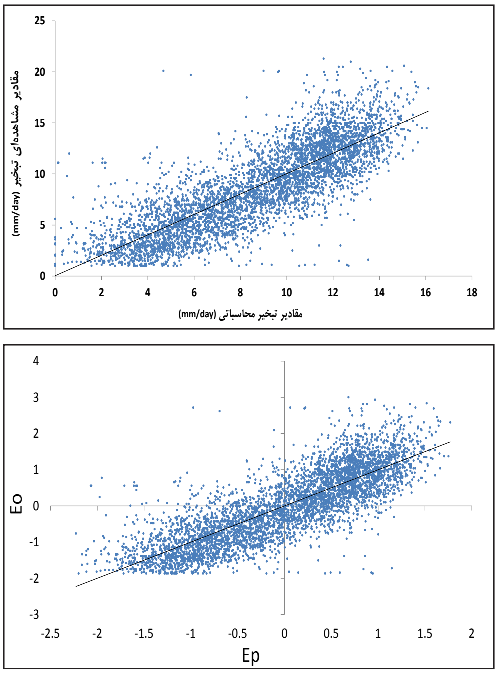
جدول ۵- برآوردهای مبتنی بر رگرسیون تیغه‌ای در مدل شماره ۵ با سه متغیر (RH_{max}, T_{min}, n)

$B^*_3(k)$	$B^*_2(k)$	$B^*_1(k)$	k
-۴۵/۱۵	۴۷/۹۰	۱۷۵/۴۱	۰
-۴۹/۸۶	۴۸/۰۳	۱۶۴/۱۱	۰/۰۵
-۵۳/۰۳	۴۸/۱۲	۱۵۴/۷۶	۰/۱
-۵۵/۱۵	۴۸/۱۳	۱۴۶/۸۳	۰/۱۵
-۵۶/۵۶	۴۸/۰۷	۱۳۹/۹۸	۰/۲
-۵۷/۴۵	۴۷/۹۳	۱۳۳/۹۷	۰/۲۵
-۵۷/۹۸	۴۷/۷۲	۱۲۸/۶۳	۰/۳
-۵۸/۲۲	۴۷/۴۵	۱۲۳/۸۴	۰/۳۵
-۵۸/۲۶	۴۷/۱۵	۱۱۹/۵۰	۰/۴
-۵۸/۱۴	۴۶/۸۰	۱۱۵/۵۴	۰/۴۵
-۵۷/۹۰	۴۶/۴۲	۱۱۱/۹۰	۰/۵
-۵۷/۱۷	۴۵/۵۹	۱۰۵/۴۴	۰/۶
-۵۶/۲۱	۴۴/۷۱	۹۹/۸۴	۰/۷
-۵۵/۱۳	۴۳/۷۹	۹۴/۹۱	۰/۸
-۵۳/۹۸	۴۲/۸۶	۹۰/۵۲	۰/۹
-۵۳/۴۰	۴۲/۳۹	۸۸/۵۰	۰/۹۵



شکل ۲- مسیر تیغه در مدل رگرسیونی با سه متغیر مستقل شامل (RHmax,Tmin, n)

بعد از محاسبات انجام شده مشخص شد که مدل منتخب با متغیرهای دمای حداقل هوا، رطوبت نسبی حداکثر و ساعات آفتابی دارای هم‌خطی توام نمی‌باشد، زیرا مدل بدست آمده با روش رگرسیون تیغه‌ای (ریچ) همان R^2 بدست آمده در حالت رگرسیون چندگانه یعنی ۰/۶۷ را بدست می‌دهد. خطای مدل رگرسیونی نیز برابر با RMSE = ۲/۳۷ بود که منطبق بر مدل منتخب ما بود. شکل (۳) دیاگرام نقاط پراکنش مقادیر تبخیر مشاهداتی را در ازای مقادیر حاصل از مدل‌های رگرسیونی خطی چندگانه (مدل منتخب شماره ۳) نشان می‌دهد. نمودار سمت راست مربوط به داده‌های غیراستاندارد و نمودار سمت چپ مربوط به داده‌های استاندارد می‌باشد. بطوریکه از این شکل می‌توان استنباط کرد ابر نقاط در اطراف خط ۱:۱ قرار دارد. از طرفی تعداد محدودی نقطه در نزدیکی محور افقی در مدل دیده می‌شود که مقدار تبخیر محاسباتی بسیار زیاد در مقابل مقدار تبخیر مشاهده‌ای (نزدیک صفر) قرار گرفته‌اند. به نظر می‌رسد این نقاط مربوط به خطای اندازه‌گیری و ثبت داده‌ها می‌باشند. Malik و Kumar (۲۰۱۵) در مطالعه خود بر روی تبخیر روزانه تشت در پانتناگار (هند)، RMSE روش MLR را در حدود ۱/۶ بدست آوردند. همچنین سیفی و همکاران (۱۳۸۹) و شیخ الاسلامی و همکاران (۱۳۹۳) در مطالعات خود که به ترتیب در ایستگاه کرمان و ایستگاه مشهد انجام داده‌اند، RMSE را برای داده‌های استاندارد شده به ترتیب ۰/۱۸۳ و ۰/۰۸ بدست آورده‌اند. Almedeij (۲۰۱۲)، Ladlani و همکاران (۲۰۱۲) و Shirsath و Kumar (۲۰۰۹) نیز به ترتیب در مطالعات خود با روش MLR میزان RMSE را ۱/۸۳، ۱ و ۱/۶ بدست آورده‌اند که این اعداد و ارقام نشانگر کارایی خوب روش رگرسیون چندگانه خطی در مناطق مختلف اقلیمی می‌باشد.



شکل ۳- نمودارهای نقاط پراکنش مقادیر تبخیر مشاهداتی به ازای مقادیر حاصل از مدل رگرسیونی خطی منتخب شماره ۳

نتیجه‌گیری

با توجه به اهمیت برآورد تبخیر در برنامه‌ریزی‌های توسعه آبیاری، در این مطالعه، امکان برآورد E (مقادیر تبخیر روزانه) با استفاده از مدل‌های رگرسیون خطی چندگانه و رگرسیون ریچ و اهمیت نسبی متغیرهای متفاوت موثر بر E مورد ارزیابی قرار گرفت. بدین منظور از داده‌های ایستگاه سینوپتیك مراغه در فاصله سال‌های ۱۳۹۱-۱۳۷۱ و ۸ متغیر هواشناسی (متغیرهای مورد استفاده در این تحقیق تبخیر از تشت (E)، دمای حداکثر (Tmax) و دمای حداقل (Tmin)، رطوبت نسبی حداکثر (RHmax) و رطوبت نسبی حداقل (RHmin)، ساعات آفتابی (n) و سرعت باد (wind)) استفاده گردید. با توجه به مشکل همبستگی بین متغیرهای اولیه (هم‌خطی توام)، در این پژوهش برای اولین بار از روش رگرسیون ریچ به عنوان پردازشگر داده‌های ورودی به مدل‌های رگرسیونی خطی چندگانه استفاده شد تا هم‌خطی توام بین متغیرهای مستقل ورودی به مدل تحلیل شوند. طبق نتایج بدست آمده برای مدل منتخب که دارای سه مولفه رطوبت نسبی حداکثر، دمای حداقل هوا و ساعات آفتابی بود، هم‌خطی توام مشاهده نشد. پس از ایجاد مدل‌های رگرسیونی خطی که بصورت نزولی از اهمیت متغیرها ایجاد شد، یک مدل از روش خطی $f(RH_{max}, T_{min}, n)$، به عنوان مدل منتخب انتخاب شد که دارای RMSE=۲/۳۷ بود. مدل‌های ایجاد شده در این مطالعه دقت مطلوبی در برآورد E داشته و قابلیت توسعه برای سایر نقاط کشور را دارا می‌باشد.

کاربرد این روش برای سایر نقاط کشور قابلیت توسعه دارد؛ بدین صورت که با داشتن آمارهای هواشناسی مرتبط با تبخیر می‌توان روابط رگرسیونی متناسب هر منطقه را بدست آورد. البته با توجه به اینکه هدف از توسعه این گونه مدل‌ها بدست آوردن تبخیر و یا هر پارامتر دیگر در مناطق فاقد آمار و یا با آمار کم است، پیشنهاد می‌شود به نوع آمار موجود در منطقه و امکان فراهم بودن آنها توجه شود. به عنوان مثال دما از جمله پارامترهایی است که در ایستگاه‌های زیادی اندازه‌گیری می‌شود

و می‌تواند در توسعه روابط کاربردی مورد استفاده قرار گیرد. همچنین می‌توان در صورت کمبود آمار هواشناسی از اطلاعات جغرافیایی و منطقه‌ای نظیر طول و عرض جغرافیایی، ارتفاع منطقه و نظیر این آمار در بدست آوردن مدل برآورد تبخیر استفاده کرد. در صورت نبود هیچگونه آمار در یک منطقه باید به دنبال منطقه‌ای مشابه آن از لحاظ اقلیمی و ارتفاعی و جغرافیایی بود تا مدل تبخیر را در آن بدست آورده و به منطقه بدون آمار تعمیم داد.

منابع

اسکافی نوغانی، م.، مفتاح هلقی، م. و مساعدی ا. ۱۳۸۷. ارائه مدل رگرسیونی پیش‌بینی تلفات تبخیر بااستفاده از پارامترهای اندازه گیری شده هواشناسی. سومین کنفرانس مدیریت منابع آب ایران، ۲۳ تا ۲۵ مهر، دانشگاه تبریز، دانشکده مهندسی عمران، تبریز.

خداشناس، س.ر.، قهرمان، ب.، داوری، ك. و ناظریان، ح ۱۳۸۷. ارائه مدل‌های رگرسیونی چند متغیره برآورد رسوب درحوضه‌های شمال استان خراسان. مجله آب و خاك، ۲: ۱۶۵-۱۵۰.

سیفی، ا.، میرلطیفی، س.م. و ریاحی، ح. ۱۳۸۹. توسعه مدل ترکیبی رگرسیون چندگانه-تحلیل مولفه‌ها و عامل‌های اصلی (PCA-MLR) در پیش بینی تبخیر-تعرق مرجع. نشریه آب و خاک، ۲۴(۶): ۱۱۸۶-۱۱۹۶.

شیخ الاسلامی، ن.، قهرمان، ب.، مساعدی، ا.، داوری، ک. و مهاجرپور م. ۱۳۹۳. پیش بینی تبخیر و تعرق گیاه مرجع (ETO) با استفاده از روش آنالیز مؤلفه های اصلی (PCA) و توسعه مدل رگرسیونی خطی چندگانه (PCA_MLR). نشریه آب و خاک (علوم و صنایع کشاورزی)، ۲۸(۲): ۴۲۰-۴۲۹.

فکری، ن.، اسماعیلی، ح.، دوستی، ح. و پارسایی، ر. ۱۳۹۰. بررسی کاربرد رگرسیون ریچ در علوم پزشکی. نشریه دانشجویی آمار (ندا)، سال یازدهم. شماره اول: ۱۲-۱۹.

نوری، ر.ا.، اشرفی، خ. و اژدرپور، ا. ۱۳۸۷. مقایسه کاربرد روش‌های شبکه عصبی مصنوعی و رگرسیون خطی چندمتغیره بر اساس تحلیل مولفه‌های اصلی برای پیش‌بینی غلظت میانگین روزانه کربن مونوکسید: بررسی موردی شهر تهران. مجله فیزیك زمین و فضا، ۳۴: ۱۳۵-۱۵۲.

Almedeij J. ۲۰۱۲. Modeling Pan Evaporation for Kuwait by Multiple Linear Regression. The Scientific World Journal. 9: 10-11.

پی‌نوشت

1- Multiple Linear Regression
2- Variance Inflation Factor
3- Correlation coefficient
4- Artificial Neural Network
5- Stephen
6- Stewart
7- Ridge Regression

Balan B, S Mohaghegh, and Ameri S. 1995. State- of-Art- in permeability determination from well log data: Part 1- A comparative study. Model development. SPE. 30978:17-25.

Bowker A.H, and Lieberman G.J. 1972. Engineering Statistics. Prentice-Hall.

Bruton JM, Mcclendon RW, and Hoogenboom G. 2000. Estimating daily pan evaporation with artificial neural networks. Trans ASAE.43(2):491–496.

Chow V. T. Maidment D. R. and Mays L.W. 1988. Applied Hydrology. McGraw hill, Newyork.

Googhari SK. 2012. Daily pan evaporation estimation using a neuro-fuzzy-based model. J Agric Sci Tech. 2: 223–228.

Hirsch R.M, Helsel D.R, Cohn T.A, Gilroy E.J. 1993. Statistical Analysis of Hydrologic Data. In: Handbook of Hydrology. Chapter 17. Maidement, R. Mc GRAW-Hill, INC.

Kisi O. 2009. Modeling monthly evaporation using two different neural computing techniques. Irrig Sci. 27:417–430.

Ladlani I, Hauichi L, Dhemili L, Heddem S, and Blouze Kh. 2012. Estimation of Daily Refrence Evapotranspiration in the North of Algeria using Adaptive Neuro-Fuzzy Inference System (ANFIS) and Multiple Linear Regression (MLR) Models: A Comparative Study. Arab J Sci Eng.39: 5959-5969.

Li, Y.F. Min, X. and Thong-ngee, G. 2010. Adaptive ridge regression system for software cost estimating on multi-collinear datasets. The Journal of System and Software. 83: 2332-2343.

Malik A, Kumar A. 2015. Pan Evaporation Simulation Based on Daily Meteorological Data Using Soft Computing Techniques and Multiple Linear Regression. Water Resour Manage. 29: 1859-1872 .

Mallik A.P, Jyothy S.A, and Sekhar Reddy K.C. 2013. Daily Reference Evapotranspiration Estimation using Linear Regression and ANN Models. The Institution of Engineers (India). 93(4):215–221.

Salas, J.D. 1993. Analysis and modeling of hydrological time series, Handbook of Hydrology, edited by David R. Maidment, McGraw-Hill, New York.

Shirgure PS, and Rajput GS. 2012. Prediction of daily pan evaporation using neural networks models. Sci J Agric. 1(5):126–137.

Shirgure PS. 2011. Evaporation modeling with neural networks-A research review. Int J Res Rev Soft Intell Comput. 1(2):37–47.

Shirsath P.b, and Kumar A.S. 2009 . A Comparative Study of Daily Pan Evaporation Estimation Using ANN, Regression and Climate Based Models.Water Resour Manage. 24: 1571-1581.

Shirsath PB, Singh AK. 2010. A comparative study of daily pan evaporation estimation using ANN, regression and climate based models. Water Resour Manag. 24:1571–1581.