ارزیابی کاربرد مدل رگرسیون چندمتغیره تیغه‌ای در برآورد تبخیر از تشت

مجید جعفری[1*] و یعقوب دین پژوه[2]

۱ *- نویسنده مسئول، دانشجوی کارشناسی ارشد منابع آب، دانشکده کشاورزی، دانشگاه تبریز. M.jafari.twone@gmail.com
۲- دانشیار گروه مهندسی آب، دانشکده کشاورزی، دانشگاه تبریز.

تاریخ دریافت: ۹۴/۷/۲۱ تاریخ پذیرش: ۹۴/۱۰/۹

چکیده

یکی از پارامترهای مهم در برنامه‌ریزی و بهره‌برداری از مخازن و طراحی سامانه‌های آبیاری، تبخیر و تعرق می‌باشد. اهمیت کاربردی تخمین هرچه دقیق تبخیر از یک سو و پیچیدگی رابطه علت و معلولی پدیده از سوی دیگر، ضرورت استفاده از روش‌های جدید داده‌کاوی را نشان می‌دهد. در این مطالعه امکان شبیه‌سازی تبخیر از تشت در ایستگاه تبریز با استفاده از مدل‌های رگرسیون چندمتغیره بررسی شد. داده‌های هواشناسی شامل حداکثر و حداقل دمای هوا، نقطه شبنم، حداکثر و حداقل رطوبت نسبی، تعداد ساعات آفتابی و سرعت روزانه باد سال‌های ۹۱-۱۳۷۱ ایستگاه سینوپتیک تبریز استفاده شدند. مدل‌های مختلف رگرسیونی چندگانه خطی و غیر خطی برای ایستگاه تبریز بسط داده شدند. مدل منتخب رگرسیون چندگانه خطی توسط روش رگرسیون تیغه‌ای مورد آزمون قرار گرفت تا چند همخطی بین ورودی‌ها در مدل لحاظ شود. مقادیر ضریب تورم واریانس برای یکایک متغیرها محاسبه شد. نتایج نشان داد که همه ضریب تورم واریانس‌ها دارای مقدار کمتر از ۱۰ بودند. افزون بر این، نسبت $\lambda_{max} / \lambda_{min}$ برای مدل منتخب با دو متغیر حداقل دمای هوا و تعداد ساعات آفتابی، برابر با ۳/۳٤ به‌دست آمد. بنابراین چند همخطی در مدل رگرسیون چندگانه خطی، منتخب مشاهده نشد. آماره دوربین واتسون نیز برای مدل منتخب برابر ۱/٤٥ به‌دست آمد که نشانگر قابل اعتماد بودن مدل رگرسیونی چندگانه خطی منتخب می‌باشد. مقادیر RMSE و R^2 برای مدل‌های منتخب به‌ترتیب برابر ۲/٤٥ میلی‌متر بر روز و ۰/٦۷ برای رگرسیون چندگانه خطی و ۲/۵۸ میلی‌متر بر روز و ۰/٦۵ برای رگرسیون چندگانه غیرخطی به‌دست آمد که این نتیجه نشانگر توانایی روش‌های رگرسیونی در تخمین تبخیر از تشت در ایستگاه تبریز می‌باشند.

کلید واژه‌ها: تبریز، تبخیر، چند همخطی، رگرسیون تیغه ای، مدل‌های رگرسیونی.

Evaluation of Multiple Ridge Regression Model to Estimation of Pan Evaporation

M. Jafari[1*]and Y. Dinpasho[2]

1*- Corresponding Author, Msc Student, Department of Water Engineering, Faculty of Agriculture, University of Tabriz, Iran.
2 – Associate Professor, Department of Water Engineering, Faculty of Agriculture, University of Tabriz, Iran.

Received:13 October 2015 Accepted:30 December 2015

Abstract

Evapotranspiration is one of the most important parameters in the Planning and operation of reservoirs, designing of irrigation systems. The practical importance of accurate estimates of evaporation and the complexity of effect phenomenon, shows the use of new methods of data mining. In this study, the simulation of pan evaporation in Tabriz station using multiple regression models were investigated. Meteorological data, including maximum and minimum air temperature, dew point, maximum and minimum air relative humidity, number of sunshine hours and Daily wind speed during (1992-2012) were used in synoptic Tabriz stations. Various models of multiple linear regression and nonlinear one were derived for Tabriz station. The selected multiple linear regression model were tested by Ridge Regression method to be considered multi-collinearity among inputs in the model.Variance inflation factor, values for each variable were calculated. The results showed that all Variance inflation factor ,s had the value less than 10. In

Archive of SID

www.SID.ir

addition, the ratio $\lambda_{max} / \lambda_{min}$ for two- variable selected model was 3.34. Therefore, there was no multi-collinearity in the selected multiple linear regression model f (Tmin, n). Durbin-Watson statistic for the selected model was 1.45 that shows the reliability of the selected multiple linear regression model. RMSE and R^2 values of the selected models (multiple linear regression and Non- Linear Regression) was calculated as 2.45 and 0.67 and 2.58 and 0.65, respectively. This results demonstrate the ability of regression techniques to estimate Pan evaporation in Tabriz station.

Keywords: Tabriz, Evaporation, Multi-collinearity, Ridge regression, Regression models.

مقدمه

تبخیر یکی از مولفه‌های اصلی چرخه‌ی هیدرولوژی و توازن آب در طبیعت و زیست‌بوم‌های کشاورزی است. همواره مفاهیم آماری به عنوان یک اصل زیربنایی در تحلیل بسیاری از مسایل پیچیده تصمیم‌گیری کاملا پذیرفته شده است. یکی از روش‌های آماری، روش‌های رگرسیونی چند متغیره خطی[۱] و غیر خطی[۲] است .هدف این‌گونه مسایل در شکل کلی آن پیدا کردن ارتباط بین متغیر وابسته و متغیرهای مستقل 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 (چاو و همکاران[۳]، ۱۹۸۸). مطالعاتی توسط بروتون و همکاران[۴] (۲۰۰۰)، شیرسات و سینگ[۵] (۲۰۱۰)، شیرگور[۶] (۲۰۱۱)، گوغاری[۷] (۲۰۱۲) و شیرگوری و راجپوت[۸] (۲۰۱۲) با استفاده از روش‌های شبکه عصبی مصنوعی و رگرسیون چندگانه خطی انجام گرفت که نشان دهنده نتایج قابل قبول برای هر دو مدل و البته نتایج قابل اعتمادتر در مدل شبکه عصبی مصنوعی می‌باشد. لی و همکاران[۹] (۲۰۱۰) سیستمی با دو سری داده متفاوت برای مقایسه دو روش رگرسیون خطی ساده و رگرسیون خطی تیغه¬ای ارائه دادند. نتایج نشان دهنده عملکرد بهتر رگرسیون تیغه¬ای در مدل‌سازی با داده‌های دارای چند همخطی نسبت به رگرسیون خطی ساده است. شیرسات و کومار[۱۰] (۲۰۰۹) تبخیر روزانه از تشت را با روش‌های شبکه‌های عصبی مصنوعی[۱۱]، رگرسیون چندگانه خطی و مدل‌های پنمن، استفن[۱۲] و استووارت[۱۳] پیش‌بینی کردند. نتایج نشان داد که روش‌های شبکه عصبی مصنوعی و رگرسیون چندگانه خطی در مقایسه با سایر روش‌ها کارایی بهتری داشتند. آلمیدیج[۱۴] (۲۰۱۲) پیش بینی 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 بازه زمانی ۱۷ ساله (از ژانویه ۱۹۹۳ تا دسامبر ۲۰۰۹) در یک منطقه بیابانی بکار برده شدند. درنهایت رابطه بین تبخیر با دما و رطوبت نسبی هوا بصورت خطی ایجاد شد. نتایج نشان داد که مدل ایجاد شده با این پارامترها همبستگی بسیار بالایی با داده‌های مشاهداتی دارد. لدلنی و همکاران[۱۵] (۲۰۱۲) تبخیر- تعرق روزانه در منطقه مدیترانه‌ای کشور الجزایر را توسط روش‌های سیستم فازی[۱۶] و رگرسیون چندگانه خطی مدل‌سازی کردند. داده‌های به‌کار رفته شامل میانگین رطوبت نسبی روزانه، ساعات آفتابی، حداکثر، حداقل و میانگین دمای هوا و سرعت باد بودند. نتایج به‌دست آمده نشان داد که عملکرد هر دو مدل در پیش بینی تبخیر – تعرق قابل قبول است. مالیک و همکاران[۱۷] (۲۰۱۳) و کیشی[۱۸] (۲۰۰۹)، برای برآورد تبخیر از تشت از روش‌های شبکه عصبی مصنوعی و رگرسیون چندگانه خطی استفاده کردند. نتایج این دو مطالعه نشان دهنده کارایی خوب مدل رگرسیون چندگانه خطی بوده است. مالیک و کومار[۱۹] (۲۰۱۵) تبخیر روزانه از تشت را توسط روش-های شبکه عصبی مصنوعی و رگرسیون چندگانه خطی در منطقه پانتاگار (هند) شبیه‌سازی کردند. مقایسه‌های انجام شده توسط معیارهای ارزیابی R^2 و RMSE نشان دهنده کارایی بهتر روش‌های شبکه عصبی مصنوعی و رگرسیون چندگانه خطی بود. اسکافی نوغانی و همکاران (۱۳۸۷) از روش رگرسیون چندگانه خطی با استفاده از پارامترهای هواشناسی، مقدار تبخیر را برآورد و با مقادیر اندازه‌گیری شده از تشتک تبخیر مقایسه کردند و منطقه مورد مطالعه ایشان حوضه گرگان‌رود بود. نتایج نشان داد که خروجی روش رگرسیون چندگانه خطی در این حوضه با مقادیر اندازه گیری شده تقریبا برابر بوده و بنابراین مدل دقت مناسب تبخیر را تخمین زد.

1 - Multiple Linear Regression (MLR)
2 - Non- Linear Regression (NLR)
3 -Chow *et al.*
4 -Bruton *et al.*
5 -Shirsath and Singh
6 -Shirgure
7 -Googhari
8 -Shirgure and Rajput
9 -Li *et al.*
10 -Shirsath and Kumar
11 -Artificial Neural Network
12 -Stephen
13 -Stewart
14 -Almedeij
15 -Ladlani *et al.*
16 -Fuzzy
17 -Malik *et al.*
18 -Kisi
19 -Malik and Kumar

Archive of SID

www.SID.ir

شکل۱- موقعیت جغرافیایی ایستگاه تبریز

علامت ستاره موقعیت ایستگاه تبریز را نشان می‌دهد.

شادمانی و معروفی (۱۳۸۸) تبخیر روزانه از تشت را در منطقه کرمان با رگرسیون چندگانه غیرخطی، شبکه عصبی مصنوعی و سیستم استنتاج فازی مدل‌سازی کردند. نتایج بدست آمده نشان از عملکرد خوب مدل رگرسیون چندگانه غیرخطی در برآورد تبخیر منطقه می‌باشد. روش رگرسیون تیغه‌ای[1] یکی از قدرتمندترین روش‌ها برای لحاظ نمودن همبستگی بین متغیرهای مختلف پیش‌بینی کننده در مدل‌سازی رگرسیونی می‌باشد. این روش تاکنون بطور گسترده در مطالعات هیدرولوژی (در ایران) به‌کار نرفته است. فکری و همکاران (۱۳۹۰) رگرسیون تیغه‌ای را به‌عنوان یکی از کاراترین روش‌های مقابله با مشکل همخطی[2] در داده‌های مربوط به عوامل خطر بیماری‌های قلبی معرفی کردند. همچنین طبری و همکاران[3] (۲۰۱۰) ، با استفاده از شبکه عصبی مصنوعی و رگرسیون چندگانه غیرخطی تبخیر روزانه از تشت را در استان همدان تخمین زدند. آنها از داده‌های بارندگی، دمای هوا، رطوبت نسبی، تابش خورشیدی و سرعت باد به عنوان ورودی مدل استفاده نمودند. نتایج نشان داد که میزان تبخیر برآورد شده به روش شبکه عصبی مصنوعی دارای خطای کمتر در مقایسه با مدل رگرسیون چندگانه غیرخطی می‌باشد و متغیرهای دما و سرعت باد، بیشترین تأثیر را در برآورد تبخیر از تشت دارند. در مطالعه حاضر، از مدل رگرسیون خطی چندگانه، رگرسیون غیرخطی و رگرسیون تیغه‌ای برای تخمین تبخیر از تشت استفاده شده است. بنابراین، هدف این مطالعه ارزیابی روش‌های رگرسیون چندمتغیره خطی، غیر خطی و رگرسیون تیغه ای برای پیش‌بینی تبخیر از تشت ایستگاه تبریز می‌باشد.

مواد و روش‌ها

تبریز مرکز استان آذربایجان شرقی به‌عنوان منطقه مورد مطالعه انتخاب شده است. این استان از سمت شمال به جمهوری‌های آذربایجان و ارمنستان، از غرب و جنوب غرب به استان آذربایجان غربی، از شرق به استان اردبیل و از جنوب شرق به استان زنجان محدود شده است. این شهرستان دارای آب و هوای سرد کوهستانی است و کلّ محدوده‌ی استان را کوه‌ها و ارتفاعات تشکیل داده‌اند. مساحت استان آذربایجان شرقی حدود ۴۵,۴۹۱ کیلومتر مربع است که از این جهت، یازدهمین استان بزرگ ایران محسوب می‌شود. این استان محل اتصال دو رشته کوه اصلی البرز و زاگرس است و بلندترین نقطه‌ی آن، قله‌ی کوه سهند است. موقعیت جغرافیایی تبریز در طول جغرافیایی ۴۶ درجه و ۱۷ دقیقه شرقی و عرض جغرافیایی ۳۷ درجه و ۵ دقیقه شمالی و ارتفاع ۱۳۶۴ متری از سطح دریا واقع شده است. دشت تبریز، با مساحت سه هزار هکتاری در استان آذربایجان شرقی بزرگترین دشت موجود در این استان (بعد از دشت‌های مرند و جلفا) می‌باشد. این دشت از شمال به دامنه‌های جنوبی کوه‌های میشو، از جنوب به دامنه‌های شمالی سهند، از شرق به محدوده شهر تبریز و از غرب به اراضی شوره‌زار مایان و دریاچه ارومیه محدود می‌شود. میزان نزولات جوی در استان به‌طور متوسط از۲۵۰ الی ۶۰۰ میلی‌متر در سال در نوسان و تبخیر از تشت نیز حدود ۱۷۰۰ میلی‌متر در سال است. فصول زمستان و بهار، فصول بارندگی منطقه محسوب می‌شوند و بیشترین شدت بارندگی در بهار مشاهده می‌شود. فصل پاییز نیز در رده سوم (از لحاظ بارندگی) پس از بهار و زمستان قرار دارد. شکل (۱) موقعیت جغرافیایی شهرستان تبریز را نشان می‌دهد.

1 -Ridge Regression
2 -Multicolinearity
3 -Tabari *et al.*

Archive of SID
www.SID.ir

داده های 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) و دمای نقطه شبنم (T_{dew}) استفاده شده است. داده‌ها از سازمان آب منطقه‌ای استان آذربایجان شرقی اخذ شده و پس از بررسی اشتراک زمانی داده‌های ثبت شده برای پارامترهای ذکر شده در بالا، سال ۱۳۷۱ به‌عنوان مبدا مطالعات انتخاب شد. به‌دلیل فقدان یا گم شدن برخی داده‌ها، از روش رگرسیون خطی ساده ما بین سری زمانی داده‌های تبریز و ایستگاه‌های همسایه برای تخمین این داده‌ها استفاده شد. بعد از تشکیل ماتریس همبستگی، معنی‌دار بودن ضرایب همبستگی بین ایستگاه‌های مختلف توسط آزمون t در سطح معنی‌داری پنج درصد مورد آزمون واقع شد. سپس مناسب‌ترین همبستگی بین داده‌های ایستگاه تبریز و سایر ایستگاه‌ها در نظر گرفته شد و داده-های گمشده تخمین زده شد. آماره آزمون t برای بررسی معنی‌دار بودن همبستگی بین متغیرها)، به شرح زیر است (میدمنت[1]، 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 می‌گردید. بعد از جاگذاری داده‌های گم‌شده (برای پارامترهای دما و رطوبت نسبی هوا)، کیفیت داده‌ها کنترل شد. برای این منظور در طول کل دوره آماری نمودار دمای حداکثر و دمای حداقل در یک نمودار و همینطور رطوبت نسبی حداکثر و رطوبت نسبی حداقل در نموداری دیگر رسم شد. در هر نمودار بیشتر بودن سری نظیر حداکثر نسبت به سری حداقل کنترل می‌شد، به‌عبارت 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 بعنوان متغیر وابسته و مقادیر متغیرهای دمای حداکثر و دمای حداقل، رطوبت نسبی حداکثر و رطوبت نسبی حداقل، ساعات آفتابی، سرعت باد و دمای نقطه شبنم بعنوان متغیرهای مستقل در نظر گرفته شدند. برای اینکه داده‌ها دارای واحد و دامنه تعریف متفاوتی بودند 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

ارتباط خطی بین متغیرهای مستقل $x_1, x_2, \ldots, x_n$ و متغیر Y در مدل رگرسیون چندگانه خطی، به‌شرح زیر در نظر گرفته شد:

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

که در آن، a_0 عرض از مبدا و $a_1, a_2, \ldots, a_n$ ضرایب رگرسیونی می‌باشند (بالان و همکاران[2]، ۱۹۹۵).

مدل رگرسیون چندگانه خطی در حالت ماتریسی به‌شرح زیر نوشته شد:

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

که در آن **β**: بردار ضرایب رگرسیون، **ε**: ماتریس خطای برازش، **Y**: بردار نظیر متغیر وابسته و **X**: ماتریس متغیرهای مستقل می‌باشند که به‌شرح زیر تعریف شدند:

$$\mathbf{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ \cdot \\ \cdot \\ \cdot \\ Y_n \end{bmatrix}$$

$$\mathbf{X} = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdot & \cdot & \cdot & x_{1p} \\ 1 & x_{21} & x_{22} & \cdot & \cdot & \cdot & x_{2p} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 1 & x_{n1} & x_{n2} & \cdot & \cdot & \cdot & x_{np} \end{bmatrix}$$

1 - Maidment

2 -Balan *et al.*

Archive of SID

www.SID.ir

$$\boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ . \\ . \\ . \\ \beta_p \end{bmatrix} \quad \boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ . \\ . \\ . \\ \varepsilon_n \end{bmatrix} \qquad (۵)$$

با حل رابطه بالا بر حسب $\boldsymbol{\beta}$ 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 رگرسیون چندگانه خطی هم برای متغیرهای مشاهداتی و هم برای متغیرهای استانداردشده بسط داده شد. افزون بر این، برای لحاظ کردن اثر وابستگی متغیرهای مستقل با همدیگر (چند همخطی)، مدل رگرسیون تیغه‌ای نیز به‌کار برده شد.

چند همخطی

واقعیت این است که بین متغیرهای مستقل ممکن است وابستگی (همبستگی معنی‌دار) وجود داشته باشد. در این حالت، استفاده از رگرسیون چندگانه خطی بدون درنظر گرفتن اثر مذکور (چند همخطی) می‌تواند گمراه کننده باشد. برای پی بردن به وجود حالت چند همخطی دو روش به شرح زیر استفاده شد:

۱. استفاده از ضریب تورم واریانس[1]، که برای مدل با p متغیر بشرح زیر محاسبه گردید:

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

که در آن R^2_j: ضریب تعیین ناشی از رگرسیون چندگانه خطی با در نظر گرفتن x_j بعنوان متغیر وابسته و p-1: متغیر دیگر به‌عنوان متغیرهای مستقل می‌باشند. به تعداد متغیرها (P) ضریب تورم واریانس محاسبه گردید و در حقیقت ضرایب تورم واریانس‌ها عناصر قطر اصلی وارون ماتریس همبستگی متغیرهای پیش‌بینی کننده هستند. در این مطالعه هر متغیری که ضریب تورم واریانس آن بیشتر از ۱۰ به‌دست می‌آمد، به‌عنوان متغیری که باعث چند همخطی می‌شود درنظر گرفته می‌شد (باوکر و لیبرمن[2]، ۱۹۷۲).

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

$$\left|\mathbf{X}'\mathbf{X}-\boldsymbol{\lambda}\mathbf{I}\right| = \mathbf{0} \qquad (۸)$$

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

رگرسیون تیغه‌ای

برای لحاظ کردن اثر چند همخطی در مدل‌سازی رگرسیونی روش‌های مختلف پیشنهاد شده است. در این مطالعه به جای استفاده از روش حداقل مربعات، برای تخمین ضرایب مدل از روش رگرسیون تیغه‌ای استفاده شد. در این مدل برای برآورد ضرایب مدل رگرسیون تیغه‌ای رابطه زیر در نظر گرفته شد (لی و همکاران، ۲۰۱۰):

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

که در آن، I: ماتریس یکه به ابعاد p*p می‌باشد. به‌طور کلی، برای هر مساله یک مقدار k بهینه وجود دارد، که ساده ترین روش یافتن آن حل رابطه (۹) به ازای چند مقدار k واقع در فاصله $0 \le k \le 1$ است. در این مطالعه مقادیر ۰/۰۵، ۰/۱۰، الی ۱، برای k در نظر گرفته شد و سپس نمودار مقادیر $\mathbf{B}^*(k)$ به‌صورت تابعی از k در یک نمودار رسم شد. این نمودار، مسیر تیغه نامیده می‌شود. مقدار k چنان برگزیده شد که برآوردهای نسبتا پایدار برای پارامترها به‌دست آید. به‌عبارت بهتر، مقادیر ضرایب با افزایش k تغییر چندانی نداشته باشد و نمودار ضرایب تقریبا افقی باشد (لی و همکاران، ۲۰۱۰). نحوه استانداردسازی

1 - Variance inflation factor

2 -Bowker and Lieberman

Archive of SID

www.SID.ir

داده‌ها در روش رگرسیونی تیغه‌ای نیز به‌شرح زیر در نظر گرفته شد:

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

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

که در آن n: تعداد داده‌ها، p: تعداد متغیرها و X_{ij}: داده i ام در متغیر j ام می‌باشد (باوکر و لیبرمن، ۱۹۷۲).

دراین مطالعه ابتدا تعداد پنج ماه از سال (از اول آبان تا آخر اسفند) به‌علت یخبندان یا عدم ثبت داده‌های نظیر تبخیر در تمام سال‌های مربوط به دوره آماری حذف گردید و سایر داده‌های تبخیر از تشت به‌صورت روزانه (از اول فروردین سال ۱۳۷۱ تا آخر مهر همان سال) در صفحه گسترده اکسل چیده شد. به‌همین ترتیب، برای سال‌های دیگر این کار در ادامه ستون همان صفحه اکسل انجام شد. بنابراین، تعداد ۲۰ سال داده روزانه (که در هر سال فقط هفت ماه– از اول فروردین تا آخر مهر– داده داشتند) در یک ستون قرار گرفتند. برای هرروز مقادیر سایر داده‌های مربوط به متغیرهای هواشناسی (هرمتغیر در یک ستون) مشابه داده‌های تبخیر در ستون‌های دیگر همان صفحه اکسل درج گردید. سپس مدل رگرسیون چندگانه خطی بین تبخیر و متغیرهای هواشناسی (مندرج در جدول ۱) در دو حالت زیر انجام شد. در حالت اول متغیرهای مستقل و وابسته بدون استاندارد شدن وارد مدل گردید و ضرایب پارامترهای مدل و ضریب تعیین به‌دست آمد. در حالت دوم، متغیرهای مذکور توسط رابطه (۱)، استاندارد و سپس وارد مدل شد و ضرایب پارامترهای مدل و ضریب تعیین به‌دست آمد. آنگاه بهترین مدل از روی ارزیابی عملکرد آن انتخاب شد. به‌طوری‌که اشاره شد، در این مطالعه علاوه بر رگرسیون خطی چندگانه از رگرسیون غیرخطی نیز استفاده شد که در ادامه شرح داده شده است.

رگرسیون غیرخطی

در این مطالعه با توجه به غیر خطی بودن فرایند تبخیر (لیو و همکاران[1]، ۲۰۰۳) از رگرسیون غیر خطی نیز استفاده شد. به عبارت، دیگر هنگامی که معادله رگرسیونی شامل عبارات غیرخطی مانند توابع نمایی[2]، لگاریتمی[3] و یا توانی[4] باشند، مدل‌های رگرسیونی مذکور، غیرخطی تلقی می‌شوند (شرستا و کازاما[5]، ۲۰۰۷). در این مدل همانند مدل رگرسیون چندگانه خطی، داده‌های تبخیر روزانه از تشت به‌عنوان متغیر وابسته و دمای حداکثر هوا، حداکثر رطوبت نسبی هوا و حداقل رطوبت نسبی هوا، میانگین سرعت باد روزانه و تعداد ساعات آفتابی به عنوان متغیرهای مستقل در نظر گرفته شدند. در این مدل رابطه‌ی رگرسیونی غیرخطی براساس کلیه پارامترهای ذکرشده در بالا به‌صورت زیر در نظر گرفته شد:

$$E = b_0.(\mathrm{T}_{\max})^{b1}.(RH_{\max})^{b2}.(RH_{\min})^{b3}.(wind)^{b4}.(n)^{b5} \qquad (۱۲)$$

که در آن، b_0، b_1، ... و b_5: ضرایب مدل رگرسیون غیر خطی هستند. برای تخمین ضرایب مذکور ابتدا از طرفین رابطه فوق به‌شرح زیر لگاریتم گرفته شد:

$$\log(E) = \log(\mathrm{b}_0) + b_1 \log(\mathrm{T}_{\max}) + b_2 \log(RH_{\max}) + b_3 \log(RH_{\min}) + b_4 \log(wind) + b_5 \log(n) \qquad (۱۳)$$

به این ترتیب، رابطه غیرخطی (۱۲) به رابطه خطی (۱۳) تبدیل شد. سپس برای تخمین پارامترهای مدل اخیر سری زمانی نظیر log E بعنوان متغیر وابسته در نظر گرفته شد و سری‌های زمانی log (T_{max})، log (RH_{max})، log (RH_{min})، log (wind)و log (n) بعنوان متغیرهای مستقل منظور گردید. آنگاه همانند مدل خطی پارامترهای مدل تخمین زده شدند. پس از محاسبه پارامتر عرض از مبدا یعنی log (b_0)، آنتی لگاریتم آن به‌عنوان ضریب b_0 در رابطه (۱۲) در نظر گرفته شد و سایر ضرایب بدون تغییر در رابطه (۱۲) منظور شدند. در این مطالعه، برای مدل غیرخطی از متغیرهای دمای حداقل هوا و دمای نقطه شبنم استفاده نشد، زیرا این سری‌ها داده‌های منفی زیادی داشته و بنابراین، قابل لگاریتم‌گیری نبودند. همچنین در برخی موارد، داده‌های متغیر ساعات آفتابی و دمای حداکثر و سرعت باد دارای مقدار صفر بودند، که در این‌گونه موارد برای ساعات آفتابی و دمای حداکثر هوا عدد ۰/۱ و برای سری زمانی سرعت باد به همه داده‌ها عدد ۰/۵ اضافه گردید. افزون بر این، چون با استاندارد کردن متغیرها ارقام منفی برای هرمتغیر ظاهر می‌گردد، بنابراین، قابل لگاریتم‌گیری نبوده و درنتیجه در مدل غیرخطی، فقط از داده‌های ثبت شده (بدون استانداردسازی) استفاده شد.

1 -Liu *et al.*
2 - Exponential Functions
3 - Logarithmic Functions
4 - Power Functions
5 -Shrestha and Kazama

Archive of SID
www.SID.ir

معیارهای ارزیابی عملکرد مدل‌ها

به منظور مقایسه و ارزیابی عملکرد مدل‌های مورد بررسی از آماره‌های میانگین مربعات خطا (RMSE)، خطای مطلق میانگین (MAE)، ضریب تعیین (R^2) و ضریب تعیین اصلاح شده ($\hat{R}^2_p$) استفاده گردید. فرمول‌های استفاده شده برای این آماره‌ها بصورت زیر بود:

$$RMSE = (\frac{1}{n}\sum_{i=1}^{n}(p_i - o_i)^2)^{0.5} \quad (۱۴)$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|p_i - o_i| \quad (۱۵)$$

$$R^2 = \frac{SS_R}{S_{yy}} \quad (۱۶)$$

$$SS_R = b'\mathbf{X}'\mathbf{Y} - \frac{1}{n}(\sum_{i=1}^{n}Y_i)^2 \quad (۱۷)$$

$$S_{yy} = \sum_{i=1}^{n}(Y_i - \bar{Y})^2 \quad (۱۸)$$

$$\bar{R}^2_p = 1 - \frac{n-1}{n-p}(1-R^2) \quad (۱۹)$$

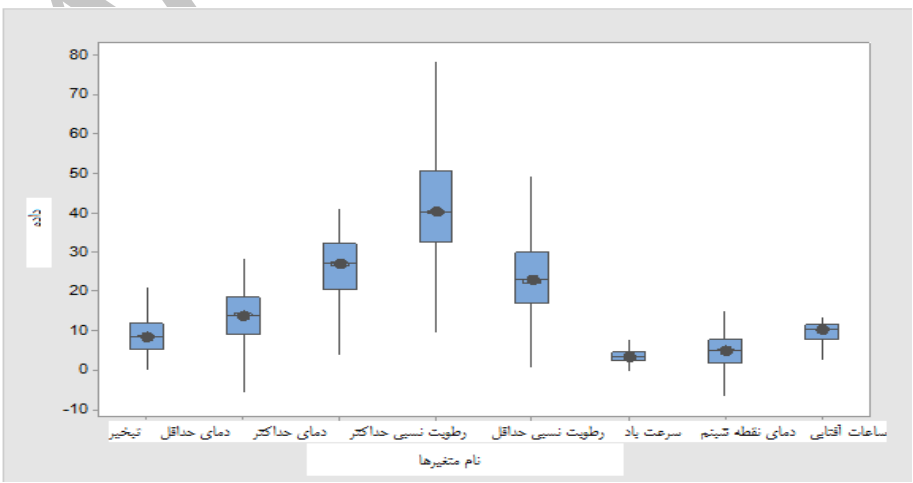
در این روابط، n: تعداد داده‌ها، O_i: مقادیر مشاهداتی تبخیر از تشت، P_i: مقادیر محاسباتی تبخیر از مدل و p: تعداد متغیرهای مستقل و SS_R: مجموع مربعات رگرسیون و S_{yy}: مجموع کل مربعات می‌باشند. بعد میانگین مربعات خطا و خطای مطلق میانگین مشابه با بعد متغیرهای مورد استفاده و ضریب تعیین بدون بعد است. هرچه ضریب تعیین به عدد یک نزدیک شود عملکرد مدل خوب ارزیابی می‌گردد. در این حالت، معمولا مقادیر آماره‌های میانگین مربعات خطا و خطای مطلق میانگین به سمت صفر میل می‌کنند. برای تشخیص خودهمبستگی سری زمانی باقیمانده‌های مدل از آماره دوربین- واتسون یا (D) به‌شرح زیر استفاده شد (باوکر و لیبرمن، ۱۹۷۲):

$$D = \frac{\sum_{t=2}^{n}(e_t - e_{t-1})^2}{\sum_{t=1}^{n}e_t^2} \quad (۲۰)$$

که در آن، e_t: در واقع t امین مقدار باقیمانده مدل می‌باشد. مقادیر بحرانی آماره دوربین- واتسون برای تعداد داده بیشتر از ۱۰۰ در سطح یک درصد بین ۱/۴۴ تا ۱/۶۵ معنی‌دار می‌باشد. در این مطالعه اگر مقدار آماره دوربین- واتسون محاسبه شده ما بین این دو عدد قرار می‌گرفت فرض وجود خودهمبستگی در خطاها رد می‌شد و مدل رگرسیونی خطی بدست آمده مناسب تلقی می‌گردید.

نتایج و بحث

شکل (۲) نمودار جعبه‌ای متغیرهای مورد استفاده در مدل‌های مختلف را در ایستگاه تبریز نشان می‌دهد. این نمودار اطلاعاتی در مورد داده‌های موجود برای هر متغیر بشرح میانه داده‌ها، چارک اول و سوم داده‌ها، واریانس داده‌های بین چارک اول و سوم که طول مستطیل‌ها می‌باشد، واریانس داده‌های بین چارک سوم و داده حداکثر و همچنین واریانس بین چارک اول و داده حداقل که به‌ترتیب با اندازه طول خطوط بالا و پایین مستطیل‌ها مشخص شده است. جدول (۱) حاوی برخی از این اطلاعات می‌باشد.



شکل ۲- نمودار جعبه‌ای متغیرهای بکار برده شده برای ایستگاه تبریز در مطالعه حاضر (۹۲- ۱۳۷۱).

جدول ۱- اطلاعات نمودار جعبه‌ای متغیرهای بکار رفته برای ایستگاه تبریز در مطالعه حاضر (۹۱- ۱۳۷۱).

Archive of SID

www.SID.ir

متغیر ها	حداکثر	حداقل	چارک اول	چارک سوم	میانه
تبخیر (میلی‌متر بر روز)	۲۲/۱۰	۰/۱۰	۵/۲	۱۱/۹۰	۸/۶
دمای حداقل هوا (درجه سانتی‌گراد)	۲۸/۲۰	-۵/۴۰	۹	۱۸/۶۰	۱۴
دمای حداکثر هوا (درجه سانتی‌گراد)	۴۱	-۱	۲۰/۶	۳۲	۲۷/۱۵
رطوبت نسبی حداکثر (درصد)	۹۵/۱۳	۹/۸۸	۳۲/۵	۵۰/۷۵	۴۰/۷۵
رطوبت نسبی حداقل (درصد)	۹۰	۱	۱۷	۳۰	۲۳
سرعت باد (متر بر ثانیه)	۱۱/۲۵	۰	۲/۵	۴/۶۲	۳/۵
دمای نقطه شبنم (درجه سانتی‌گراد)	۱۵/۰۶	-۱۲/۸۸	۲	۷/۷۷	۵/۰۷۵
ساعات آفتابی (ساعت)	۱۳/۵۰	۰	۸	۱۱/۷	۱۰/۴

جدول ۲- ماتریس متقارن ضرایب همبستگی بین متغیرهای هواشناسی مورد استفاده در ایستگاه تبریز

نام متغیر	دمای حداکثر	دمای حداقل	حداکثر رطوبت نسبی	حداقل رطوبت نسبی	تعداد ساعات آفتابی	سرعت باد	دمای نقطه شبنم
دمای حداکثر	۱	۰/۹۳۰	-۰/۷۶۷	-۰/۷۰۳	۰/۶۱۸	۰/۲۲۲	۰/۴۹۳
دمای حداقل	۰/۹۳۰	۱	-۰/۶۲۹	-۰/۵۲۵	۰/۵۴۷	۰/۳۵۴	۰/۶۰۷
حداکثر رطوبت نسبی	-۰/۷۶۷	-۰/۶۲۹	۱	۰/۹۱۱	-۰/۵۸۵	-۰/۱۴۲	۰/۱۳۵
حداقل رطوبت نسبی	-۰/۷۰۳	-۰/۵۲۵	۰/۹۱۱	۱	-۰/۵۳۶	-۰/۰۹۱	۰/۱۴۴
تعداد ساعات آفتابی	۰/۶۱۸	۰/۵۴۷	-۰/۵۸۵	-۰/۵۳۶	۱	۰/۱۴۳	۰/۲۰۵
سرعت باد	۰/۲۲۲	۰/۳۵۴	-۰/۱۴۲	-۰/۰۹۱	۰/۱۴۳	۱	۰/۲۲۹
دمای نقطه شبنم	۰/۴۹۳	۰/۶۰۷	۰/۱۳۵	۰/۱۴۴	۰/۲۰۵	۰/۲۳	۱

جدول (۲) مقادیر ماتریس همبستگی بین متغیرهای ورودی به مدل‌های مختلف را نشان می‌دهد. به‌طوری که از جدول (۲) می‌توان استنباط کرد، بیشترین ضریب همبستگی برای متغیر حداکثر دمای هوای تبریز متعلق به متغیر نظیر حداقل دمای هوای همان ایستگاه (معادل ۰/۹۳) می‌باشد که باتوجه به تعداد ۴۷۷۵ زوج داده‌های استفاده شده این رقم در سطح پنج درصد معنی‌دار می‌باشد. پس از متغیر دمای حداقل هوا، می‌توان ادعا کرد که متغیر حداکثر رطوبت نسبی هوا بیشترین ضریب همبستگی منفی (۰/۷۷-) با متغیر حداکثر دمای هوای تبریز را دارد. بنابراین، می‌توان نتیجه گرفت که با افزایش درجه حرارت هوا مقدار رطوبت نسبی آن کاهش پیدا می‌کند. گرچه کمترین مقدار ضریب همبستگی بین دمای حداکثر و متغیرهای هواشناسی متعلق به پارامتر سرعت باد (۰/۲۲) می‌باشد، و لیکن همین عدد نیز در سطح ۵٪ معنی‌دار می‌باشد. بیشترین و کمترین مقدار ضریب همبستگی برای متغیر رطوبت نسبی حداکثر بترتیب متعلق به پارامترهای رطوبت نسبی حداقل و دمای نقطه شبنم معادل با ۰/۹۱ و ۰/۱۳۵ می‌باشد که هردو درسطح ۵٪ معنی‌دار بودند. همچنین بیشترین و کمترین مقدار ضریب همبستگی برای متغیر ساعات آفتابی به‌ترتیب متعلق به پارامترهای دمای حداکثر و سرعت باد معادل با ۰/۶۲ و ۰/۱۴ می‌باشد که هردو در سطح پنج درصد معنی‌دار بودند. همانطور که از جدول (۲) مشاهده می‌شود کمترین همبستگی پارامترهای مختلف هواشناسی تبریز مربوط به متغیر سرعت باد می‌باشد، بطوریکه، بیشترین مقدار ضریب همبستگی بین سرعت باد و دمای حداقل (معادل ۰/۳۵) و کمترین آن متعلق به همبستگی سرعت باد و رطوبت نسبی حداقل (معادل ۰/۰۹-) بود. بنابراین می‌توان دریافت که سرعت باد همبستگی نسبتا کمتری با متغیرهای هواشناسی (در مقایسه با دیگر متغیرها) دارد. در مدل‌های استفاده شده احتمال ایجاد داده منفی در خروجی وجود دارد. لذا به‌دلیل اینکه تبخیر نمی‌تواند مقدار منفی بخود بگیرد، در نتیجه بجای داده‌های منفی عدد صفر (در خروجی) منظور شد.

جدول (۳) نتایج حاصل از اجرای مدل‌های منتخب رگرسیون چندمتغیره خطی، غیرخطی و تیغه‌ای را نشان می‌دهد. به‌طوریکه از این جدول می‌توان استنباط کرد، در این جدول، متغیرهای مورد نیاز برای مدل‌های منتخب، شامل دمای حداکثر و حداقل هوا، رطوبت حداکثر و حداقل هوا، سرعت باد و ساعات آفتابی می‌باشد که عرض از مبدا و ضرایب b_1 تا b_6 هر مدل به‌همراه آماره‌های عملکرد در ستون‌های دوم تا دوازدهم جدول نشان داده شده‌اند. مقادیر آماره ضریب تعیین در مدل‌های مختلف، از ۰/۶۵۴ تا ۰/۶۷۲ تغییر می‌کنند. افزون بر این، مقادیر ضریب تعیین اصلاح شده نیز بین ۰/۶۵۴ تا ۰/۶۷۲ تغییر می‌کنند. ضمنا آماره دوربین واتسون مدل‌ها از ۱/۴۴ تا ۱/۵۹ تغییر می‌کند. مقادیر میانگین مربعات خطا نیز از ۲/۴۵ تا ۲/۵۸ میلی‌متر بر روز تغییر می‌کنند. ملاحظه شد که با کاهش تعداد متغیرهای ورودی برای مدل‌ها در هر دو روش رگرسیون خطی و غیرخطی میزان خطا افزایش و میزان ضریب تعیین و ضریب تعیین اصلاح شده کاهش می‌یابد،

Archive of SID

www.SID.ir

ولی این افزایش خطا یا کاهش ضریب تعیین باتوجه به ارقام موجود در ستون‌های میانگین مربعات خطا، خطای مطلق میانگین و ضریب تعیین در هرکدام از مدل‌های تکراری رگرسیون چندمتغیره خطی و غیرخطی، ناچیز و در حدود پنج الی ۱۰ درصد بین اولین مدل با بیشترین متغیر و آخرین مدل با فقط یک متغیر می‌باشد. لازم به ذکر است مدل رگرسیون چندگانه خطی دارای شش مدل با تعداد متغیر پسرو از شش تا یک، مدل رگرسیون غیرخطی با پنج مدل با تعداد متغیر پسرو از پنج تا یک و مدل تیغه‌ای دارای پنج مدل بود که از میان آنها بهترین مدل برای هرکدام انتخاب و در جدول (۳) ارائه گردید. پر واضح است که هرچه مدل ساده‌تر باشد، نسبت به مدل‌های پیچیده‌تر (با تعداد متغیر زیاد) بهتر خواهد بود. با بررسی های انجام گرفته برای رگرسیون خطی، غیرخطی و تیغه‌ای، مدل‌های منتخب هرکدام با دو متغیر مورد نیاز برای محاسبه تبخیر از تشت به‌عنوان مدل مناسب تشخیص داده شدند. متغیرهای مورد نیاز برای مدل‌های خطی و تیغه‌ای شامل دمای حداقل هوا و ساعات آفتابی، و برای مدل رگرسیون غیرخطی این متغیرها شامل دمای حداکثر و سرعت باد بود.

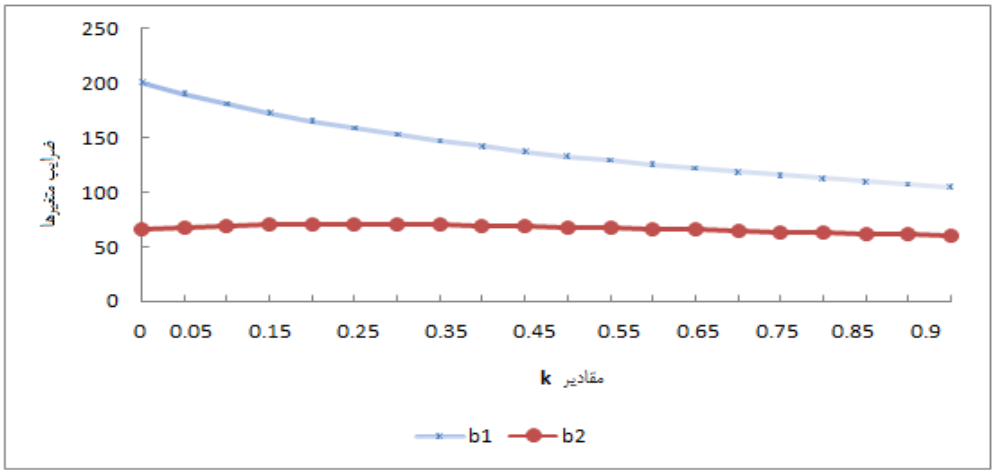
برای پی بردن به وجود حالت چند همخطی از مقادیر ویژه ماتریس همبستگی ($\boldsymbol{\lambda}$) و عناصر روی قطر اصلی وارون ماتریس همبستگی $(\mathbf{X'X})^{-1}$ در مدل منتخب استفاده شد. در این مطالعه ماتریس $(\mathbf{X'X})$ و معکوس آن یعنی $(\mathbf{X'X})^{-1}$ در مدل منتخب رگرسیون خطی چندگانه به شرح زیر حاصل شد:

$$\mathbf{X'X}=\begin{bmatrix}1 & 0.538\\ 0.538 & 1\end{bmatrix},\quad (\mathbf{X'X})^{-1}=\begin{bmatrix}1.41 & -0.76\\ -0.76 & 1.41\end{bmatrix}$$

مقادیر ویژه ماتریس $(\mathbf{X'X})$ با استفاده از نرم افزار Maple به شرح λ_1 = ۱/۵۳۸ و λ_2 = ۰/۴۶ به‌دست آمد و بنابراین، نسبت $\lambda_{max}/\lambda_{min}$ برابر با ۳/۳۴ حاصل شد باتوجه به اینکه نسبت مذکور کمتر از ۱۰ می‌باشد. بنابراین، می‌توان نتیجه گرفت که احتمال وجود چند همخطی در پارامترهای مربوط به مدل رگرسیونی منتخب خیلی ناچیز می‌باشد. در جدول (۴) مقادیر ضرایب رگرسیون تیغه‌ای به ازای مقادیر مختلف k ارائه شده است. شکل (۳) نیز مسیر تیغه‌ای ضرایب رگرسیونی b_1 و b_2 را به ازای مقادیر مختلف k نشان می‌دهد. به‌طوری که از شکل (۳) می‌توان استنباط کرد به ازای مقدار k معادل با ۰/۰۵ مسیر تیغه ها تا حدودی به‌هم نزدیک و افقی شده است. با استفاده از رقم مذکور و نیز b_0=۸/۶۴ مدل رگرسیون تیغه‌ای به شرح زیر حاصل شد:

$$E = 8.64 + 190.116\,T^{*}_{min} + 68.162\,n^{*} \qquad (۲۱)$$

که در آن، T^{*}_{min} و n^{*} : به‌ترتیب مقادیر دمای حداقل هوا و تعداد ساعات آفتابی هوا در شکل استاندارد شده با فرمول شماره دهم و E مقدار تبخیر بر حسب میلی‌متر بر روز می‌باشند. مثلا اگر دمای حداقل هوا ۱۳/۳ درجه سانتی‌گراد و تعداد ساعات آفتابی ۵/۲ باشد، با توجه به میانگین و انحراف معیار این دو پارامتر، معادل استاندارد آنها به‌ترتیب برابر ۰/۰۰۰۷۶- و ۰/۰۱۸۶۸- می‌باشد که در این‌صورت مقدار تبخیر برابر ۷/۲۲ میلی‌متر بر روز به‌دست آمد و مقدار تبخیر مشاهده شده در همان روز معادل ۷/۷ میلی‌متر بر روز است.



شکل۳- مسیر تیغه در مدل رگرسیونی با دو متغیر مستقل شامل (دمای حداقل و ساعات آفتابی).

Archive of SID

www.SID.ir

جدول ۳- ضرایب و آماره‌های ارزیابی عملکرد مدل‌های منتخب رگرسیون چندمتغیره خطی، غیرخطی و تیغه‌ای

مدل ها	b_0	b_1	b_2	b_3	b_4	b_5	b_6	میانگین مربعات خطا	ضریب تعیین	دوربین- واتسون	خطای مطلق میانگین	ضریب تعیین اصلاح شده
مدل خطی	-۰/۶۴	۰/۰۰	۰/۴۸	۰/۰	۰/۰	۰/۰۰	۰/۳۰	۲/۴۵۰	۰/۶۷۰	۱/۴۵۰	۱/۸۵۰	۰/۶۶۷
مدل خطی با داده استاندارد	۰/۰۰	۰/۰۰	۰/۶۸	۰/۰	۰/۰	۰/۰۰	۰/۲۲	۰/۵۷۲	۰/۶۷۰	۱/۴۵۰	۰/۴۳۰	۰/۶۶۷
مدل غیرخطی	۰/۰۵	۱/۳۹	۰/۰۰	۰/۰	۰/۰	۰/۳۷	۰/۰۰	۲/۵۸۰	۰/۶۵۴	۱/۵۹۰	۱/۵۳۰	۰/۶۵۴
مدل تیغه‌ای	۸/۶۴	۱۹۰/۱۲	۰/۰۰	۰/۰	۰/۰	۰/۰۰	۶۸/۱۶	۲/۴۶	۰/۶۷۲	۱/۴۴۰	۱/۷۸۰	۰/۶۷۲

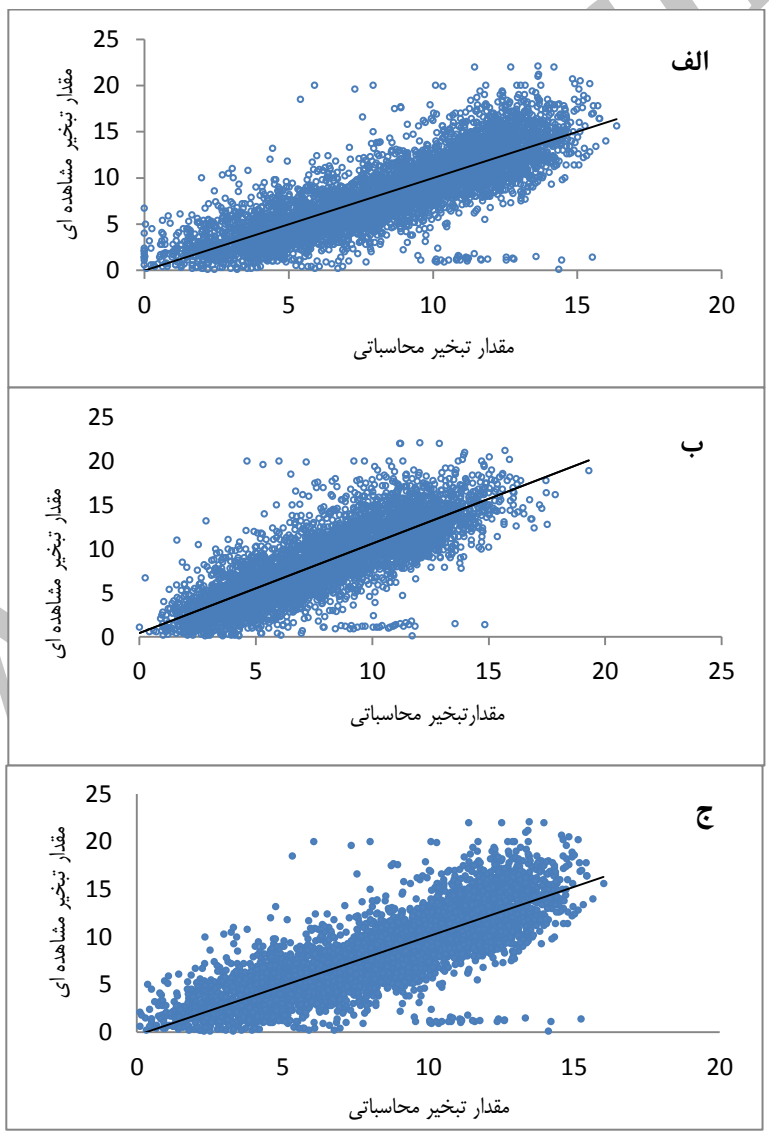
در جدول b_0 عرض از مبدا و b_1 تا b_6 بترتیب ضرایب پارامترهای (دمای حداکثر، دمای حداقل، رطوبت نسبی حداکثر، رطوبت نسبی حداقل، سرعت باد و ساعات آفتابی) می‌باشد.

Archive of SID

www.SID.ir

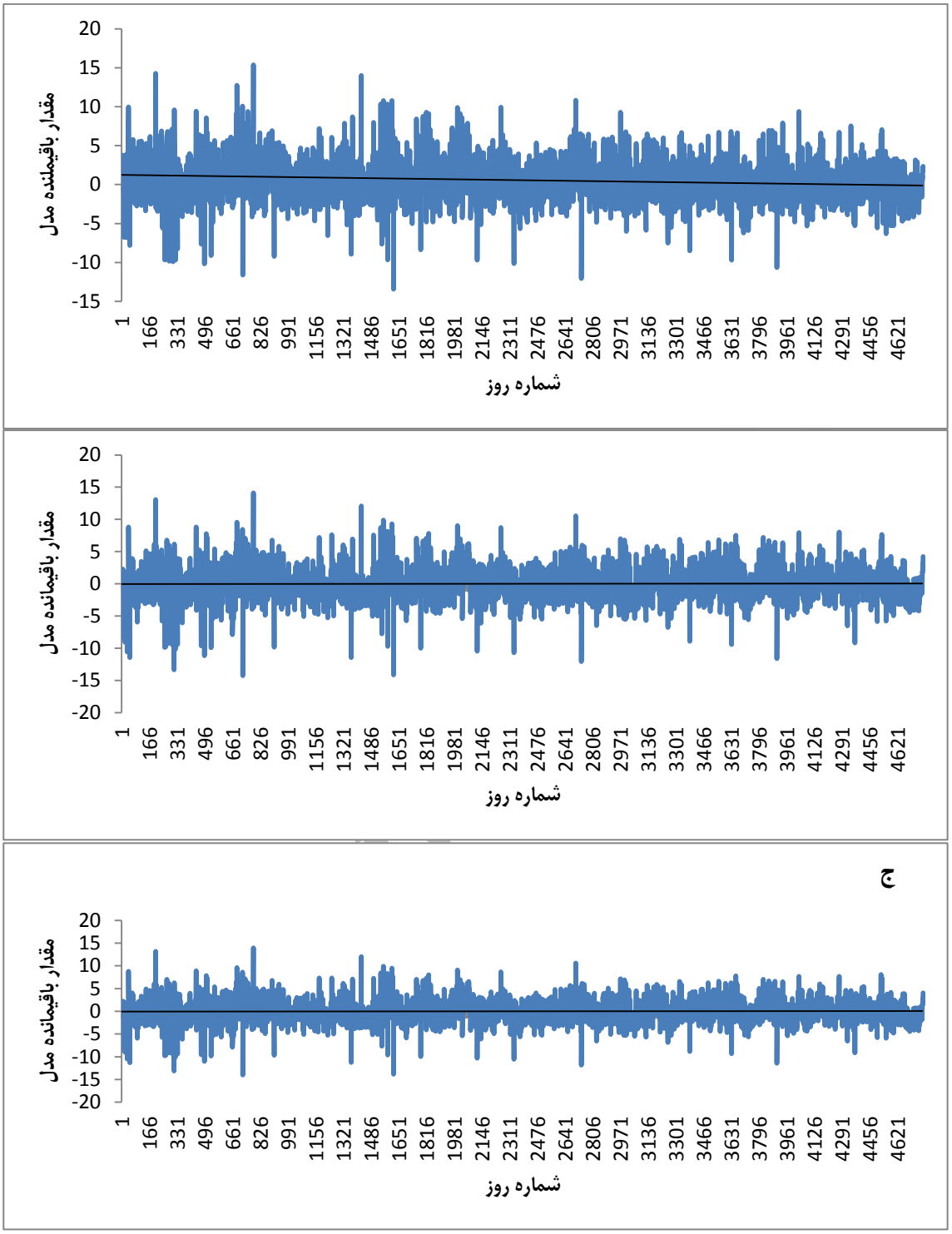
جدول٤- برآوردهای مبتنی بر رگرسیون تیغه‌ای در مدل منتخب با دو متغیرساعات آفتابی و دمای حداقل

$B^*_2(k)$	$B^*_1(k)$	k
۶۵/۷۴۸	۲۰۰/۹۲	۰/۰
۶۸/۱۶۲	۱۹۰/۱۱۶	۰/۰۵
۶۹/۶۵۴	۱۸۰/۷۴	۰/۱۰
۷۰/۴۹	۱۷۲/۴۹	۰/۱۵
۷۰/۸۵	۱۶۵/۱۴	۰/۲۰
۷۰/۸۷	۱۵۸/۵۳	۰/۲۵
۷۰/۶۳	۱۵۲/۵۳	۰/۳۰
۷۰/۲	۱۴۷/۰۵۵	۰/۳۵
۶۹/۶۳	۱۴۲/۰۲۳	۰/۴۰
۶۸/۹۵	۱۳۷/۳۷۶	۰/۴۵
۶۸/۲۰	۱۳۳/۰۶۶	۰/۵۰
۶۷/۳۹	۱۲۹/۰۵۴	۰/۵۵
۶۶/۵۵	۱۲۵/۳۰۵	۰/۶۰
۶۴/۸	۱۱۸/۴۹	۰,۷۰
۶۳	۱۱۲/۴۴	۰,۸۰
۶۱/۲۲۸	۱۰۷/۰۳	۰,۹۰



شکل٤- نمودارهای نقاط پراکنش مقادیر تبخیر مشاهده شده (میلی‌متر بر روز)، به ازای مقادیر حاصل از مدل-های منتخب، الف) رگرسیون خطی، ب) رگرسیون غیرخطی و ج) رگرسیون تیغه‌ای.

www.SID.ir



شکل۵- نمودارهای سری زمانی باقی‌مانده‌های نظیر مدل منتخب، الف) رگرسیون خطی، ب) رگرسیون غیرخطی و ج) رگرسیون تیغه‌ای. (شماره روز اول مربوط به تاریخ ۱۰ خرداد ۱۳۷۱ می‌باشد)

بعد از محاسبات انجام شده مشخص شد که مدل منتخب با متغیرهای دمای حداقل هوا و ساعات آفتابی دارای چند همخطی نمی‌باشد، زیرا مدل به‌دست آمده با روش رگرسیون تیغه‌ای همان ضریب تعیین به‌دست آمده در حالت رگرسیون چندگانه یعنی ۰/۶۷ را به‌دست می‌دهد. میانگین مربعات خطای مدل رگرسیونی نیز برابر با ۲/۴۶ میلی‌متر بر روز بود که منطبق بر مدل منتخب بود. شکل (۴) نمودار نقاط پراکنش مقادیر تبخیر مشاهده شده را در ازای مقادیر حاصل از مدل‌های رگرسیونی چند متغیره خطی،

www.SID.ir

غیرخطی و تیغه‌ای نشان می‌دهد. به‌طوری‌که از این شکل می‌توان استنباط کرد ابر نقاط در اطراف خط ۱:۱ تقریبا برای همه مدل‌هاوجود دارد. با این حال، در مدل تیغه‌ای، ابر نقاط بطور نامحسوسی با پهنای کمتر در مقایسه با مدل‌های رگرسیون چندمتغیره خطی و غیر خطی قرار گرفته است. از طرفی تعداد محدودی نقطه در نزدیکی محور افقی در تمام مدل‌ها دیده می‌شود که مقدار تبخیر محاسباتی بسیار زیاد در مقابل مقدار تبخیر مشاهده‌ای (نزدیک صفر) قرار گرفته‌اند. به‌نظر می‌رسد، این نقاط مربوط به خطای اندازه گیری و ثبت داده‌ها می‌باشند. شکل (۵) نمودار سری زمانی باقی‌مانده‌های مدل‌های مختلف رگرسیون خطی، غیرخطی و تیغه‌ای را نشان می‌دهد. می‌توان دریافت که خطای مدل‌های مختلف فاقد روند خاصی می‌باشند و مدل‌ها دارای نمودار تقریبا مشابه برای باقی‌مانده‌ها می‌باشند. ولی همان‌طور که از نمودار مدل تیغه‌ای پیداست، پراکندگی باقیمانده‌های این مدل حول محور صفر نزدیک‌تر هستند. مالیک و کومار (۲۰۱۵) در مطالعه خود بر روی تبخیر روزانه تشت در پانتناگار (هند)، میانگین مربعات خطا روش رگرسیون چندگانه خطی را در حدود ۱/۶ به‌دست آوردند. همچنین سیفی و همکاران (۱۳۸۹) و شیخ الاسلامی و همکاران (۱۳۹۳) در مطالعات خود که به‌ترتیب در ایستگاه کرمان و ایستگاه مشهد انجام داده‌اند، میانگین مربعات خطا را برای داده‌های استاندارد شده به‌ترتیب ۰/۱۸۳ و ۰/۰۸ بدست آورده‌اند. آلمیدیج (۲۰۱۲)، لدلنی و همکاران (۲۰۱۲) و شیرسات و کومار (۲۰۰۹) نیز به‌ترتیب در مطالعات خود با روش رگرسیون چندگانه خطی میزان میانگین مربعات خطا را به‌ترتیب 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.

نتیجه گیری

در مطالعه حاضر سعی شد تا مقدار تبخیر از تشت در مقیاس روزانه به‌صورت مدل‌های رگرسیون خطی، غیرخطی و تیغه‌ای در ایستگاه تبریز به تفضیل تخمین زده شود و خواننده بتواند به آسانی با روش‌های مذکور آشنایی پیدا کند. چون در اغلب نقاط صعب العبور، امکان اندازه گیری مستقیم تبخیر از تشت نیست، بنابراین با استفاده از روش‌شناسی به‌کاربرده شده می‌توان مقدار تبخیر را به آسانی از روی مدل برآورد کرد. اهمیت این کار بویژه در جاهایی که داده تبخیر وجود ندارد یا دارای آمار کافی نمی‌باشد، به‌طور قابل ملاحظه‌ای بالا می‌رود. گرچه در مطالعات قبلی مانند سیفی و همکاران (۱۳۸۹)، برای تخمین تبخیر و تعرق پتانسیل، از روش رگرسیون تیغه‌ای استفاده نشده است، ولیکن برای غلبه بر اثر همبستگی بین متغیرها، نامبردگان از روش تجزیه به مولفه‌های اصلی بهره برده‌اند. شیخ الاسلامی و همکاران (۱۳۹۳)، نیز از روش مشابه برای تخمین تبخیر و تعرق در ایستگاه مشهد، استفاده کرده‌اند. افزون بر این، کوور و ناناداگیری[1] (۲۰۰۷)، نیز برای توسعه مدل رگرسیونی برای تخمین تبخیر از تشت از روش رگرسیون خطی و تجزیه به مولفه‌های اصلی بهره برده‌اند. نامبردگان نیز برای ارزیابی کارایی مدل‌ها از آماره‌های عملکرد ضریب تعیین و مجموع مربعات خطا استفاده کرده‌اند. در مطالعه حاضر، برای اینکار از روش رگرسیون تیغه‌ای برای اولین بار در مورد تبخیر استفاده شد. در این مطالعه برای تخمین تبخیر از تشت، از داده‌های هفت متغیر هواشناسی اندازه گرفته شده در ایستگاه سینوپتیک تبریز در فاصله سال‌های ۱۳۹۱-۱۳۷۱ استفاده گردید. این متغیرها شامل مقادیر دمای حداکثر و دمای حداقل، رطوبت نسبی حداکثر و رطوبت نسبی حداقل، ساعات آفتابی، سرعت باد و دمای نقطه شبنم بود و پس از مدل‌بندی کارایی مدل‌ها با داده‌های اندازه گرفته شده از تشت تبخیر ارزیابی گردید. به‌طوری که انتظار می‌رفت متغیرهای مستقل با همدیگر وابستگی شدیدی داشتند. بنابراین لزوم حذف اثر وابستگی بین متغیرهای مستقل احساس شد و در پی اینکار از روش رگرسیون تیغه‌ای که در حذف اثر همبستگی بین متغیرها یک روش قدرتمند محسوب می‌شود استفاده شد. نتایج نشان داد که مقدار تبخیر به‌دست آمده با استفاده از مدل رگرسیون تیغه‌ای به‌طور قابل ملاحظه‌ای با مقادیر اندازه گرفته شده مشابهت دارند. در این پژوهش برای اولین بار جهت تخمین تبخیر از تشت از روش رگرسیون تیغه‌ای به‌عنوان پردازشگر داده‌های ورودی به مدل‌های رگرسیونی خطی چندگانه استفاده شد تا چند همخطی بین متغیرهای مستقل ورودی به مدل تحلیل شوند. طبق نتایج بدست آمده برای مدل منتخب که دارای دو مولفه دمای حداقل هوا و ساعات آفتابی بود، چند همخطی مشاهده نشد. پس از ایجاد مدل‌های رگرسیونی خطی و غیرخطی که بصورت نزولی از اهمیت متغیرها ایجاد شد، یک مدل از روش خطی و یک مدل از روش غیر خطی، بعنوان مدل‌های منتخب انتخاب شدند که میانگین مربعات خطا هریک به‌ترتیب ۲/۴۵ و ۲/۵۸ میلی‌متر بر روز بودند. مدل‌های ایجاد شده در این مطالعه دقت مطلوبی در برآورد تبخیر از تشت داشته و قابلیت توسعه برای سایر ایستگاه‌های کشور با اقلیم‌ها متفاوت را دارا می‌باشد. نتایج بدست آمده در این مطالعه با توجه به داده‌های اندازه گرفته شده با محدودیت طول دوره آماری به‌دست آمده است. بنابراین برای نتیجه گیری بهتر لازم است مطالعه مشابه برای داده‌های طولانی مدت انجام شود.

1 -Kovoor and Nandagiri

Archive of SID

www.SID.ir

منابع

۱- اسکافی نوغانی، م.، مفتاح هلقی، م. و ا، مساعدی. 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.

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

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

۴- شادمانی، م. و ص، معروفی. ۱۳۸۸. مقایسه چند روش برآورد تبخیر روزانه از تشت، مطالعه موردی ایستگاه کرمان. مجله علوم و فنون کشاورزی. علوم آب و خاک، سال پانزدهم. ۵۵: ۸۳-۶۹.

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

۶- فکری، ن.، اسماعیلی، ح.، دوستی، ح. و ر، پارسایی. ۱۳۹۰. بررسی کاربرد رگرسیون تیغه‌ای در علوم پزشکی. نشریه دانشجویی آمار (ندا)، سال یازدهم. ۱: ۱۹-۱۲.

۷- نوری، ر.ا.، اشرفی، خ. و ا، اژدرپور. ۱۳۸۷ .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، ۳۴: ۱۵۲-۱۳۵.

8- Almedeij, J. 2012. Modeling pan evaporation for Kuwait by multiple linear regression. The Scientific World Journal, 9: 10-11.

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

10- Bowker, H. and G. J. Lieberman. 1972. Engineering Statistics. Prentice-Hall, 852 p.

11- Bruton, J.M., Mcclendon, R.W. and G. Hoogenboom. 2000. Estimating daily pan evaporation with artificial neural networks. Transaction ASAE, 43(2):491–496.

12- Chow, V. T., Maidment, D. R. , L.W Mays. 1988. Applied hydrology. McGraw hill, Newyork, 570 p.

13- Googhari, S.K. 2012. Daily pan evaporation estimation using a neuro-fuzzy-based model. Journal of Agriculture Science and Technology, 2: 223–228.

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

15- Kovoor, G.M. and L. Nandagiri. 2007. Developing regression models for predicting pan evaporation from climate data. Journal of Irrigation and Drainage Engineering, 133(5): 444-454.

16- Ladlani, I., Hauichi, L., Dhemili, L., Heddem, S. and Kh. Blouze. 2012. Estimation of daily refrence evapotranspiration in the north of Algeria using adaptive neuro-Fuzzy inference system (ANFIS) and multiple linear regression Models: a comparative study. Arabian Journal for Science and Engineering, 39: 5959-5969.

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

Archive of SID

www.SID.ir

18- Liu, C.W., Lin, K.H., and Y.M. Kuo. 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.

19- Maidement, R.1993. Handbook of hydrology. Mc GRAW-Hill, INC, 1730p.

20- Malik, A. and A. Kumar. 2015. Pan evaporation simulation based on daily meteorological data using soft computing techniques and multiple minear megression. Water Resources Management, 29: 1859-1872.

21- Mallik, A.P., Jyothy, S.A. and K..C. Sekhar Reddy. 2013. Daily reference evapotranspiration estimation using linear regression and Ann models. The Institution of Engineers (India), 93(4):215–221.

22- Shirgure, PS. 2011. Evaporation modeling with neural networks-A research review. International Journal of Research and Reviews in Applied Sciences. 1(2):37–47.

23- Shirgure, PS. and G.S. Rajput. 2012. Prediction of daily pan evaporation using neural networks models. Science Journal Agriculcure 1(5):126–137.

24- Shirsath, P.B. and A.S. Kumar. 2009. A comparative study of daily pan evaporation estimation using Ann, regression and climate based models. Water Resources Management, 24: 1571-1581.

25- Shirsath, PB. and A.K. Singh. 2010. A comparative study of daily pan evaporation estimation using Ann, regression and climate based models. Water Resources Management, 24:1571–1581.

26- Shrestha, S. and F. Kazama. 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.

27- Tabari, H., Marofi, S. and A. A Sabziparvar. 2010. Estimation of daily pan evaporation using artificial neural network and multivariate non-linear regression. Irrigation Science, 28: 399-406.

Archive of SID

www.SID.ir