



دانشگاه گوارزی و منابع طبیعی

مجله پژوهش‌های حفاظت آب و خاک

جلد نوزدهم، شماره دوم، ۱۳۹۱

<http://jwfst.gau.ac.ir>

کاربرد مدل‌سازی فازی بر مبنای خوشه‌بندی c-mean در تخمین تبخیر از تشت (مطالعه موردی: استان خوزستان)

*سیده رویا عقیلی^۱، سعید برومندنسب^۲ و مهدی کاهه^۳

^۱دانش‌آموخته کارشناسی‌ارشد گروه آبیاری و زه‌کشی، دانشگاه شهید چمران اهواز، استاد گروه آبیاری و زه‌کشی،

دانشگاه شهید چمران اهواز، ^۲دانشجوی دکتری گروه سازه‌های آبی، دانشگاه شهید چمران اهواز

تاریخ دریافت: ۸۹/۱۱/۲۳؛ تاریخ پذیرش: ۹۰/۸/۱۰

چکیده

در این مقاله به کمک منطق فازی و خوشه‌بندی فازی به روش c-mean، ترکیب‌های مختلف با ۴ داده هواشناسی در دسترس شامل دمای هوا، رطوبت نسبی، سرعت باد و ساعات آفتابی روزانه به‌عنوان داده‌های ورودی برای پیش‌بینی تبخیر از تشت مورد بررسی قرار گرفت. برای این منظور از پارامترهای هواشناسی مربوط به سال‌های ۸۶ و ۸۷ در ۳ ایستگاه سینوپتیک اهواز، آبادان و ایذه واقع در استان خوزستان استفاده شد. نتایج به‌دست آمده، توانایی مدل استنتاج فازی تاکاجی-سوگنو و خوشه‌بندی فازی به روش c-mean را در پیش‌بینی تبخیر از تشت در ۳ ایستگاه یاد شده به‌خوبی نشان داد، به‌طوری‌که مدل‌های فازی با ۲ و ۳ خوشه، پارامتر تبخیر از تشت را با مربع ضریب همبستگی حدود ۰/۸۶-۰/۹۱ پیش‌بینی می‌کند.

واژه‌های کلیدی: تبخیر از تشت، مدل فازی تاکاجی-سوگنو، خوشه‌بندی فازی، c-mean

*مسئول مکاتبه: aghili_roya@yahoo.com

مقدمه

فرآیند تبخیر یکی از مؤلفه‌های اصلی چرخه آب در طبیعت است که نقش اساسی در مطالعات کشاورزی، هیدرولوژیکی و هواشناسی، بهره‌برداری مخازن، طراحی سیستم‌های آبیاری و زه‌کشی، زمان‌بندی آبیاری و مدیریت منابع آب ایفا می‌کند. برای اندازه‌گیری مستقیم تبخیر از سطح آزاد آب در ایستگاه‌های هواشناسی از تشت‌های تبخیر استاندارد شده استفاده می‌شود به طوری که این روش به‌عنوان یکی از روش‌های با دقت بالا برای برآورد تبخیر و تعرق گیاه مرجع شناخته شده است (رحیمی‌خوب، ۲۰۰۶). از آنجایی که راه‌اندازی و بهره‌برداری از ایستگاه‌های تبخیرسنجی در مناطق بدون مخازن آب گران تمام می‌شود و همچنین بسیاری از ایستگاه‌ها بدون آمارهای مداوم و طولانی‌مدت هستند، از این‌رو روش‌های تحلیلی و تجربی زیادی برای پیش‌بینی تبخیر از سطح آزاد آب و تبخیر از تشت با استفاده از داده‌های هواشناسی ارائه شده است. روش‌های تحلیلی به داده‌های هواشناسی زیادی نیاز دارند و بعضی از این داده‌ها مثل تابش خورشید در همه‌جا و همیشه اندازه‌گیری نمی‌شود. روش‌های تجربی به داده‌های کم‌تری نیاز دارند ولی دقت آن‌ها در مناطق مختلف متفاوت است و برای هر منطقه باید واسنجی شوند (لیناگر، ۱۹۹۴). در طی سال‌های اخیر محققان زیادی در صدد مدل کردن پدیده تبخیر برآمده اند. تبخیر از سطح آزاد آب تحت‌تأثیر پارامترهایی چون تابش خورشید، دما، رطوبت نسبی، سرعت باد و... بوده که پارامترهای اخیر نیز بر یکدیگر تأثیر متقابل می‌گذارند. در نتیجه این پدیده از یک سیستم پیچیده و غیرخطی پیروی می‌کند که تخمین دقیق آن بسیار دشوار است. از این‌رو استفاده از مدل‌های ریاضی مانند سیستم استنتاج فازی با درک رفتارهای غیرخطی سیستم برای حل این مشکل مناسب است. در مورد کاربرد منطق فازی در مدیریت منابع آب پژوهش‌های بسیاری صورت گرفته است که از این میان می‌توان به مطالعه صورت گرفته توسط ماهابیر و همکاران (۲۰۰۳) اشاره کرد. آن‌ها با استفاده از منطق فازی به پیش‌بینی رواناب در حوضه‌ای از کشور کانادا پرداختند. لوهانی و همکاران (۲۰۰۵)، با استفاده از منطق فازی مدلی را برای پیش‌بینی جریان سیلابی رودخانه نارمادا در هندوستان مرکزی طراحی کرد. آیواز و همکاران (۲۰۰۷)، با استفاده از خوشه‌بندی فازی به روش c-mean و الگوریتم ژنتیک سفره‌های آب زیرزمینی را ناحیه‌بندی و پارامترهای هیدرولیکی آن را تعیین کردند. همچنین در چند سال اخیر مدل‌های فازی و عصبی برای پیش‌بینی تبخیر از تشت براساس داده‌های هواشناسی مورد بررسی قرار گرفته است. کسکین و همکاران (۲۰۰۴)، داده‌های ایستگاه هواشناسی نزدیک دریاچه اجیردیر در غرب ترکیه را برای تعیین تبخیر روزانه از تشت با استفاده از مدل فازی مورد بررسی قرار دادند. ضریب همبستگی و مجذور

میانگین مربعات خطای مدل تدوین شده توسط ایشان به ترتیب $0/85$ و $1/6$ میلی متر در روز برآورد شد. مقدم نیا و همکاران (۲۰۰۹)، دو مدل شبکه عصبی و سیستم استنتاج فازی-عصبی تطبیقی^۱ را برای تعیین تبخیر روزانه از تشت با استفاده از داده های ایستگاه هواشناسی نزدیک مخزن چاه نیمه در سیستان شامل دمای هوا، سرعت باد، کمبود فشار بخار اشباع و رطوبت نسبی مورد مقایسه قرار دادند. همچنین کسکین و همکاران (۲۰۰۹)، مدل فازی (۲۰۰۴) خود را با مدل ANFIS مقایسه کردند. آنها نشان دادند که مدل ANFIS نسبت به مدل فازی در تخمین تبخیر روزانه دریاچه اجیردیر قابلیت بالاتری دارد. هدف از این پژوهش، نشان دادن توانمندی مدل سازی سیستم استنتاج فازی تاکاجی-سوگنو با استفاده از تکنیک خوشه بندی فازی به روش c-mean در پیش بینی پارامتر تبخیر از تشت از روی برخی داده های هواشناسی است. به کارگیری مدل هوشمند مناسب می تواند ضمن کاهش هزینه های اندازه گیری، خلا ناشی از کمبود آمارهای مداوم و طولانی مدت در برخی ایستگاه های هواشناسی را جبران نماید. در این پژوهش از داده های هواشناسی متداول مانند متوسط دمای هوا، متوسط رطوبت نسبی، سرعت باد و ساعات آفتابی که در بیش تر ایستگاه های هواشناسی قابل اندازه گیری هستند استفاده شده است.

مواد و روش ها

محدوده مطالعه و منابع داده ها: در این پژوهش، سعی شده است ایستگاه های انتخابی تا حد امکان از نظر اقلیمی و آب و هوایی با یکدیگر متفاوت باشند، از این رو ۳ ایستگاه سینوپتیک ایذه، اهواز و آبادان در استان خوزستان به ترتیب با شرایط آب و هوایی کوهستانی، خشک و مرطوب مورد بررسی قرار گرفته شده است که در شکل ۱ موقعیت این ایستگاه ها مشاهده می شود. همچنین در جدول ۱ مشخصات کلی این ایستگاه ها به همراه متوسط سالانه برخی پارامترهای هواشناسی آنها ارائه شده است. از آنجایی که آمار ورودی مدل به صورت روزانه وارد شده است در نتیجه، ۲ سال با حدود ۷۵۰ داده روزانه انتخاب گردید. داده ها مربوط به سال های ۱۳۸۶ و ۱۳۸۷ و شامل متغیرهای روزانه تبخیر از تشت^۲، متوسط دمای هوا^۳، متوسط رطوبت نسبی^۴، سرعت باد^۵ و ساعات آفتابی^۶ است.

1- Adaptive Neuro Fuzzy Inference System (ANFIS)

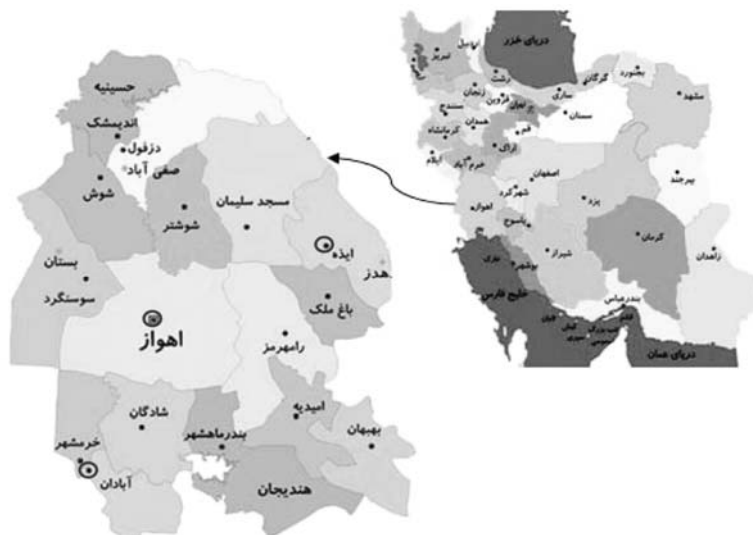
2- E_{pan}

3- T_a

4- RH

5- U

6- n



شکل ۱- موقعیت منطقه و ایستگاه‌های سینوپتیک مورد مطالعه.

جدول ۱- مشخصات کلی ایستگاه‌های سینوپتیک مورد مطالعه.

ایستگاه	طول جغرافیایی	عرض جغرافیایی	ارتفاع ایستگاه (متر)	متوسط دما (درجه سانتی‌گراد)	متوسط رطوبت نسبی (درصد)	سرعت باد (متر بر ثانیه)	تبخیر از تشت (mm/d)
اهواز	۴۸ درجه و ۴۴ دقیقه	۳۱ درجه و ۲۰ دقیقه	۲۲/۵	۲۲/۵	۴۴/۷۵	۵	۸/۴
آبادان	۴۸ درجه و ۱۲ دقیقه	۳۰ درجه و ۲۲ دقیقه	۶/۶	۲۳/۹	۷۲/۵	۳	۷/۵
ایذه	۴۹ درجه و ۵۱ دقیقه	۳۱ درجه و ۵۰ دقیقه	۸۲۷	۱۹/۸	۵۴/۸	۵/۵	۶

در این پژوهش پس از مرور انجام شده بر منابع و ارزیابی اولیه بین هر یک از پارامترهای موجود و همچنین با توجه به حجم عملیات مدل‌سازی، ۱۰ سناریو انتخاب شد (جدول ۲). به طور مثال در ارزیابی اولیه، بین پارامترهای تبخیر، سرعت باد و ساعات آفتابی، به تنهایی رابطه منطقی مشاهده نشد و این دو پارامتر به عنوان پارامتر مکمل در ترکیب‌ها قرار گرفتند. تبخیر از تشت به عنوان خروجی و بقیه متغیرها به عنوان داده‌های ورودی مدل استفاده شدند. پس از ایجاد سناریوهای مختلف حدود ۷۰ درصد داده‌های موجود به بخش واسنجی مدل‌ها و بقیه به بخش صحت‌سنجی اختصاص داده شد. همچنین در این پژوهش برای جلوگیری از منفی شدن مقادیر به دست آمده از مدل‌سازی و کاهش مقادیر خطا از لگاریتم طبیعی داده‌ها استفاده شده است.

جدول ۲- ساختارهای مختلف داده‌های ورودی مدل‌های فازی.

پارامترهای ورودی	ردیف
T_a	۱
RH	۲
n و T_a	۳
U و T_a	۴
n و RH	۵
U و RH	۶
RH و T_a	۷
n و RH, T_a	۸
U و RH, T_a	۹
U و n , RH, T_a	۱۰

خوشه‌بندی به روش **c-mean**: در این پژوهش به منظور استخراج قواعد و پارامترهای بخش فرض و نتیجه در مدل استنتاج فازی تاکاجی- سوگنو از خوشه‌بندی فازی به روش **c-mean** استفاده شده است. این روش که اولین بار توسط بزدرک (۱۹۸۱) ارائه گردید روشی کارآمد برای خوشه‌بندی فازی داده‌ها است. اگر بردار داده‌های ورودی به صورت $X = [\bar{x}_1, \dots, \bar{x}_n]$ باشد، در خوشه‌بندی فازی به این روش داده‌ها به C خوشه با درجه عضویت مشخص به هر خوشه تقسیم‌بندی می‌شوند. حداکثر تعداد خوشه‌های مورد استفاده $\sqrt{n} < c_{\max} < 2$ می‌باشد که در آن n تعداد مشاهدات آماری است. میزان درجه عضویت هر داده به هر خوشه در ماتریس عضویت $U = [u_{i,j}]_{c \times n}$ مشخص می‌شود. روش **c-mean** سعی دارد تا برای یک مجموعه نقاط، داده افرازهایی بیابد تا تابع هدف زیر را کمینه کند (کوره‌پزان‌دزفولی، ۲۰۰۸):

$$J_m(U, V) = \sum \sum u_{ij}^{m'} d_{ij}^2 \quad (1)$$

که در آن، $m' \in [1, \infty)$ است و میزان فازی بودن را مشخص می‌کند و معمولاً بین اعداد ۲-۱/۲۵ را به خود می‌گیرد. $V = [v_{ki}]_{m \times c}$ ماتریس مراکز خوشه‌ها و d_{ij} : فاصله اقلیدسی بین بردار مرکز خوشه i ام و بردار داده j ام است.

$$d_{ij} = \left[\sum_{k=1}^m (x_{kj} - v_{ki})^2 \right]^{\frac{1}{2}} \quad (2)$$

که در آن، m : تعداد متغیرها و v_{ki} : مرکز خوشه i ام از متغیر k ام است. برای کمینه کردن تابع هدف J از الگوریتم تکراری به‌عنوان جایگزینی بهینه به‌شرح زیر استفاده می‌شود:

مرحله اول: پارامترهای m' و c ، یک عدد ثابت کوچک برای ε به‌عنوان حداکثر خطای مطلوب و ماتریس عضویت U به‌طور تصادفی انتخاب می‌شود.

مرحله دوم: مرکز خوشه‌ها با استفاده از رابطه زیر به‌دست می‌آید:

$$v_{ik} = \frac{\sum_{j=1}^n u_{ij}^{m'} x_{kj}}{\sum_{j=1}^n u_{ij}} \quad (3)$$

مرحله سوم: با استفاده از رابطه زیر ماتریس عضویت جدید محاسبه می‌شود:

$$u_{ij} = \left[\sum_{i=1}^c \left(\frac{d_{ij}}{d_{kj}} \right)^{\frac{2}{m'-1}} \right]^{-1} \quad (4)$$

مرحله چهارم: این کار تا زمانی ادامه می‌یابد که اختلاف ماتریس U یا V از مراحل ۱ و ۲ با مقادیر محاسبه شده کوچک‌تر از ε شود.

یکی از مهم‌ترین مشکلات در خوشه‌بندی انتخاب تعداد خوشه مناسب است. تعداد خوشه‌ای مناسب است که در ابتدا نمونه‌های موجود در یک خوشه تا حد امکان به یکدیگر شبیه باشند و سپس نمونه‌های متعلق به خوشه‌های متفاوت با یکدیگر نامتشابه باشند. به‌عبارت دیگر خوشه‌ها باید ماکزیم فشردگی را داشته باشند و جدایی آن‌ها از یکدیگر نیز زیاد باشد. اگرچه در الگوریتم خوشه‌بندی c -mean و نسخه‌های مختلف آن تعداد خوشه‌ها باید از قبل مشخص باشد ولی در ابتدای کار تعداد مناسب آن برای طراحی مشخص نمی‌باشد و بیش‌تر با روش سعی و خطا تعیین می‌شود. برای مشخص کردن تعداد مناسب خوشه‌ها توابع ارزیابی مختلفی وجود دارد که در این پژوهش از توابع ارزیابی ارائه شده در جدول ۳ استفاده شده است و انتخاب تعداد خوشه مناسب با ماکزیم یا مینیمم کردن این توابع به‌دست می‌آید (ژانگ و همکاران، ۲۰۰۸؛ وو و یانگ، ۲۰۰۵؛ کیم و همکاران، ۲۰۰۴).

جدول ۳- الگوریتم‌های ارزیابی تعداد خوشه به کار رفته در این پژوهش.

بهترین حالت	الگوریتم	ردیف
ماکزیمم	$V_{MPC} = 1 - \frac{c}{c-1}(1 - V_{PC})$	۱
مینیمم	$V_{XB} = \frac{\sum_{i=1}^c \sum_{j=1}^n u_{ij}^m \ x_j - v_i\ ^2}{n \min_{i \neq j} \ v_i - v_j\ ^2}$	۲
مینیمم	$SC_1 = \frac{\sum_{i=1}^c \ v_i - \bar{v}\ ^2 / c}{\sum_{i=1}^c \left(\sum_{j=1}^n u_{ij}^m \ x_j - v_i\ ^2 / \sum_{j=1}^n u_{ij} \right)}$	۳

ساخت مدل فازی با استفاده از نتایج به دست آمده از خوشه بندی: سیستم استنتاج فازی یک سیستم مبتنی بر قواعد منطقی اگر-آن‌گاه است که با استفاده از قواعد مزبور می‌توان ارتباط بین تعدادی متغیر ورودی و خروجی را به دست آورد. بنابراین از یک سیستم استنتاج فازی می‌توان به‌عنوان یک مدل پیش‌بینی استفاده کرد (وحیدیان‌کامیاد و طارقیان، ۲۰۰۲). در این پژوهش پس از خوشه‌بندی فازی و انتخاب بهترین تعداد خوشه، مدل فازی تاکاجی-سوگنو برای هر یک از سناریوها ایجاد می‌شود. مراحل ایجاد بخش‌های مختلف آن براساس نتایج به دست آمده از خوشه‌بندی به شرح زیر است:

۱- تعیین پارامترهای بخش فرض و نتیجه: در این پژوهش تابع عضویت به کار رفته در بخش فرض از نوع گوسی است که با دو پارامتر $\{c, \sigma\}$ مشخص می‌شود:

$$\mu(x; c, \sigma) = e^{-\frac{1}{2} \left(\frac{x-c}{\sigma} \right)^2} \quad (5)$$

که در آن، پارامتر c : نشان‌دهنده مرکز و σ : عرض تابع عضویت گوسی است که برای به دست آوردن آن طبق روابط زیر از ماتریس عضویت به دست آمده از روش خوشه‌بندی استفاده می‌شود:

$$c_{ik} = \frac{\sum_{j=1}^n u_{ij}^m x_{kj}}{\sum_{j=1}^n u_{ij}} \quad (6)$$

$$\sigma_{ik} = \sqrt{\frac{\sum_{j=1}^n u_{ij} (x_{kj} - c_{ik})^2}{\sum_{j=1}^n u_{ij}}} \quad (7)$$

بخش نتیجه نیز در مدل استنتاج فازی تاکاجی - سوگنو یک معادله رگرسیون خطی به صورت زیر است:

$$y_i = \theta_{.i} + \theta_{1i}x_1 + \dots + \theta_{ki}x_k \quad (8)$$

که در آن، θ_i : ضرایب معادله رگرسیونی بوده و با حداقل کردن مربعات خطای وزنی میان داده‌های مشاهداتی و نتایج مدل‌سازی به روش زیر به دست می‌آید:

$$\theta_i = \{\theta_{0i}, \theta_{1i}, \dots, \theta_{ki}\} = [X_e^T W_i X_e]^{-1} X_e^T W_i Y \quad (9)$$

که در آن،

$$X_e = [1, X], \quad X = \begin{bmatrix} x_1^T \\ x_2^T \\ \vdots \\ x_j^T \end{bmatrix}, \quad Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_j \end{bmatrix}, \quad W_i = \begin{bmatrix} u_{i1} & \cdot & \dots & \cdot \\ \cdot & u_{i2} & \dots & \cdot \\ \vdots & \vdots & \ddots & \vdots \\ \cdot & \cdot & \dots & u_{ij} \end{bmatrix}$$

۲- ایجاد قوانین اگر- آن‌گاه فازی: در مدل‌سازی فازی با استفاده از خوشه‌بندی تعداد قوانین با تعداد خوشه‌های انتخاب شده در هر سناریو برابر است. شکل قوانین به کار رفته در این پژوهش با توجه به تعداد خوشه‌ها به صورت زیر تعریف شده است:

$$R_i : IF x_1 \text{ is } A_{i1} \text{ and } x_2 \text{ is } A_{i2} \dots \text{ and } x_n \text{ is } A_{in} \\ THEN y_i = \theta_{.i} + \theta_{1i}x_1 + \dots + \theta_{ki}x_k; \quad i = 1, 2, \dots, c$$

که در آن، R_i : i امین قانون فازی، A_i : تابع عضویت گوسی برای n امین پارامتر ورودی در i امین قانون فازی و y_i : خروجی i امین قانون فازی می‌باشد.

ارزیابی مدل‌های ایجاد شده با استفاده از آماره‌های مختلف و انتخاب بهترین مدل: در مدل‌سازی، همواره استفاده از مدلی ترجیح داده می‌شود که خطای کم‌تری داشته باشد. برای انتخاب بهترین مدل،

آماره‌های مختلفی مورد استفاده قرار می‌گیرد و در این پژوهش از آماره‌های مربع ضریب همبستگی^۱، مجذور میانگین مربعات خطا^۲ و میانگین درصد خطای نسبی^۳، در دو بخش واسنجی و صحت‌سنجی استفاده شد:

$$R^2 = \left(\frac{\sum_{i=1}^n (E_o - \bar{E}_o)(E_e - \bar{E}_e)}{\sqrt{\sum_{i=1}^n (E_o - \bar{E}_o)^2 \sum_{i=1}^n (E_e - \bar{E}_e)^2}} \right)^2 \quad (10)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (E_e - E_o)^2}{n}} \quad (11)$$

$$MAPE = \left(\frac{\sum_{i=1}^n \frac{|E_e - E_o|}{E_o} \times 100}{n} \right) \quad (12)$$

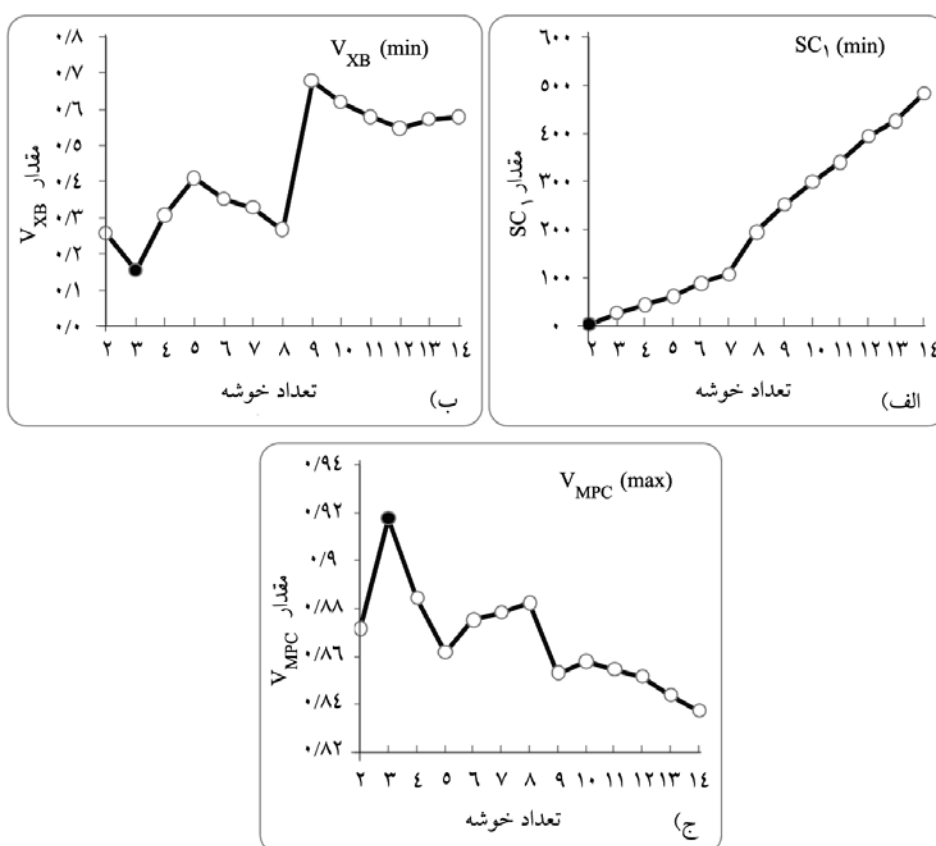
لازم به ذکر است در این پژوهش تمام الگوریتم c-mean، الگوریتم توابع ارزیابی برای تعیین تعداد خوشه مناسب و همچنین مدل فازی تاکاجی در محیط نرم‌افزار MATLAB کدنویسی و برای استفاده اصلاح شده است.

نتایج و بحث

همان‌طور که گفته شد در این پژوهش به منظور استخراج قواعد و پارامترهای مدل استنتاج تاکاجی - سوگنو از روش خوشه‌بندی فازی c-mean استفاده شد. پس از تقسیم‌بندی داده‌ها به دو بخش واسنجی و صحت‌سنجی با نسبت ۷۰ به ۳۰، داده‌های بخش آموزش به روش c-mean خوشه‌بندی شد و درجه عضویت هر داده به هر خوشه و مراکز خوشه‌ها تعیین گردید. از آنجایی که در بیش‌تر مطالعات حداکثر تعداد خوشه مناسب بین $\sqrt{n} < c_{max} < 2$ می‌باشد و تعداد داده‌های مورد استفاده در بخش آموزش برای سناریوهای مختلف حدود ۵۰۰ روز است با توجه به بررسی‌های اولیه داده‌ها بین ۱۴-۲

-
- 1- Correlation Coefficient (R^2)
 - 2- Root Mean Square Error (RMSE)
 - 3- Mean Absolute Percentage Error (MAPE)

خوشه طبقه‌بندی شد. پس از خوشه‌بندی فازی برای سناریوهای مختلف و به‌دست آوردن ماتریس عضویت فازی و مراکز خوشه‌ها، خوشه‌بندی انجام شده با استفاده از توابع ارزیابی مورد سنجش قرار گرفت. برای نمونه، شکل ۲ نحوه انتخاب بهترین تعداد خوشه برای سناریوی ۳ در ایستگاه اهواز را نشان می‌دهد، به طوری که هر یک از توابع SC_1 ، V_{XB} و V_{MPC} به ترتیب تعداد ۲، ۳ و ۳ را به‌عنوان بهترین تعداد خوشه نشان می‌دهند. خلاصه نتایج به‌دست آمده از انتخاب بهترین تعداد خوشه برای سناریوهای مختلف در جدول ۴ ارائه شده است.



شکل ۲- نحوه انتخاب بهترین تعداد خوشه با استفاده از توابع ارزیابی برای سناریوی ۳ در ایستگاه اهواز.

جدول ۴- خلاصه نتایج به دست آمده از انتخاب بهترین خوشه برای سناریوهای مختلف با استفاده از توابع ارزیابی.

ردیف	پارامترهای ورودی	بهترین تعداد خوشه		
		اهواز	آبادان	ایذه
۱	T _a	۲	۲	۲
۲	RH	۲	۲	۲
۳	n و T _a	۳، ۲	۴ و ۳، ۲	۲
۴	U و T _a	۲	۲	۳، ۲
۵	n و RH	۳، ۲	۴ و ۳، ۲	۲
۶	U و RH	۲	۲	۳، ۲
۷	RH و T _a	۲	۲	۲
۸	n و RH، T _a	۳، ۲	۳، ۲	۴ و ۳، ۲
۹	U و RH، T _a	۲	۲	۳، ۲
۱۰	U و n، RH، T _a	۳، ۲	۳، ۲	۳، ۲

همان‌طور که مشاهده می‌شود در هر سه ایستگاه مجموعه داده‌ها در بیش‌تر سناریوها به ۲ و ۳ خوشه تقسیم‌بندی می‌شوند که این تعداد می‌تواند رابطه مستقیمی با خصوصیت اقلیمی منطقه مورد مطالعه داشته باشد. به طوری که در دو ایستگاه اهواز و آبادان سناریوهایی که تنها شامل پارامترهای دما، رطوبت نسبی و سرعت باد است، تمایل به ۲ خوشه شدن را دارند و هنگامی که پارامتر ساعات آفتابی نیز به ترکیب سناریوها اضافه می‌شود داده‌ها علاوه بر ۲ خوشه به ۳ خوشه و در آبادان حتی به ۴ خوشه نیز تقسیم می‌شوند، در حالی که در ایستگاه ایذه سناریوهایی که شامل پارامترهای دما، رطوبت نسبی و ساعات آفتابی است، تمایل به ۲ خوشه شدن را دارند و هنگامی که پارامتر سرعت باد نیز به ترکیب سناریوها اضافه می‌شود داده‌ها علاوه بر ۲ خوشه به ۳ خوشه نیز تقسیم می‌شوند.

پس از خوشه‌بندی، مدل‌سازی براساس سناریوهای تعریف شده و نتایج به دست آمده از خوشه‌بندی صورت گرفت. به منظور ارزیابی کارایی هر یک از مدل‌ها و انتخاب بهترین آن‌ها از

آماره‌های R^2 ، RMSE و MAPE در دو بخش واسنجی و صحت‌سنجی استفاده شد. نتایج به‌دست آمده از مدل‌سازی در هر ۳ ایستگاه مورد مطالعه (جدول‌های ۵ تا ۷) نشان می‌دهد که با افزایش تعداد پارامترهای ورودی نتایج بهتری در مدل‌سازی ایجاد شده که این امر بیانگر اهمیت انتخاب پارامترهای ورودی در تخمین مقدار تبخیر می‌باشد.

جدول ۵- ارزیابی مدل‌های ایجاد شده با استفاده از آماره‌های مختلف در ایستگاه اهواز.

ردیف	سناریوها	تعداد خوشه مناسب	R^2		RMSE		MAPE	
			واسنجی	صحت‌سنجی	واسنجی	صحت‌سنجی	واسنجی	صحت‌سنجی
۱	T_a	۲	۰/۷۸۴	۰/۷۴۴	۲/۷۸۲	۲/۹۳۹	۴۰ درصد	۴۱ درصد
۲	RH	۲	۰/۷۵۴	۰/۷۳۶	۲/۹۷۶	۳/۰۴۳	۳۹ درصد	۴۴ درصد
۳	T_a و n	۲	۰/۷۹۰	۰/۷۴۹	۲/۷۳۹	۲/۹۳۴	۳۹ درصد	۴۱ درصد
۴	T_a و U	۳	۰/۸۰۹	۰/۷۶۹	۲/۶۱۳	۲/۷۹۹	۳۸ درصد	۳۸ درصد
۵	RH و n	۲	۰/۶۵۹	۰/۷۳۴	۳/۶۸۴	۳/۰۷۵	۳۹ درصد	۴۳ درصد
۶	RH و U	۳	۰/۷۴۸	۰/۷۴۷	۳/۰۳۵	۲/۹۹۶	۳۸ درصد	۴۳ درصد
۷	T_a و RH	۲	۰/۸۴۱	۰/۸۱۵	۲/۳۷۹	۲/۵۰۱	۳۲ درصد	۴۲ درصد
۸	T_a ، RH و n	۲	۰/۸۴۳	۰/۸۱۴	۲/۳۵۷	۲/۵۱۴	۳۲ درصد	۳۵ درصد
۹	T_a ، RH و U	۳	۰/۸۴۷	۰/۸۲۶	۲/۳۳۰	۲/۴۳۳	۳۲ درصد	۳۴ درصد
۱۰	T_a ، RH، n و U	۲	۰/۸۶۱	۰/۸۳۳	۲/۲۱۷	۲/۳۵۹	۳۰ درصد	۳۳ درصد
		۲	۰/۸۶۶	۰/۸۳۷	۲/۱۸۱	۲/۳۴۲	۳۰ درصد	۳۳ درصد
		۳	۰/۸۶۴	۰/۸۴۹	۲/۱۹۷	۲/۲۴۸	۳۰ درصد	۳۳ درصد

سیده رویا عقیلی و همکاران

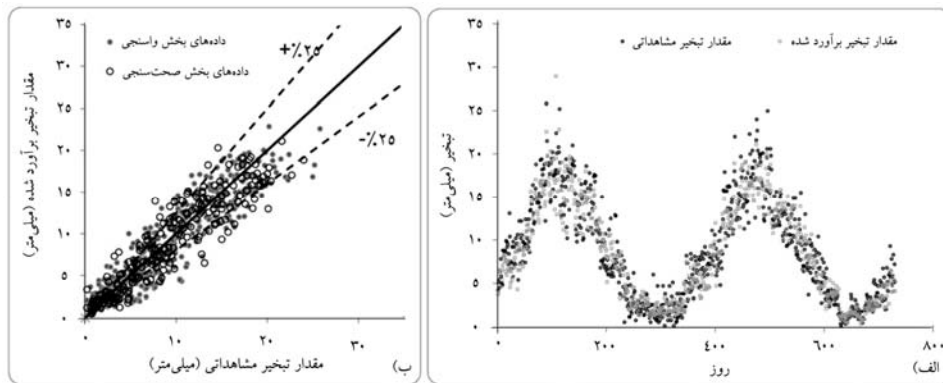
جدول ۶- ارزیابی مدل‌های ایجاد شده با استفاده از آماره‌های مختلف در ایستگاه آبادان.

ردیف	سناریوها	تعداد خوشه مناسب	R ^۲		RMSE		MAPE	
			واسنجی	صحت‌سنجی	واسنجی	صحت‌سنجی	واسنجی	صحت‌سنجی
۱	T _a	۲	۰/۷۱۴	۰/۷۳۸	۳/۷۲۱	۳/۷۵۳	۴۷ درصد	۴۰ درصد
۲	RH	۲	۰/۷۵۵	۰/۷۸۷	۳/۴۵۳	۳/۳۸۱	۶۵ درصد	۴۸ درصد
۳	n و T _a	۳	۰/۷۵۵	۰/۷۸۰	۳/۴۴۱	۳/۴۳۶	۴۲ درصد	۳۹ درصد
۴	U و T _a	۲	۰/۷۸۷	۰/۸۰۵	۳/۲۰۸	۳/۱۹۱	۴۶ درصد	۳۸ درصد
۵	n و RH	۳	۰/۷۷۴	۰/۷۹۲	۳/۳۱۶	۳/۲۹۹	۶۰ درصد	۴۷ درصد
۶	U و RH	۲	۰/۷۵۸	۰/۷۹۰	۳/۴۳۰	۳/۳۴۵	۶۵ درصد	۶۵ درصد
۷	RH و T _a	۲	۰/۸۳۳	۰/۸۴۳	۲/۸۲۶	۲/۸۶۵	۳۶ درصد	۳۴ درصد
۸	n و RH, T _a	۳	۰/۸۳۹	۰/۸۴۹	۲/۷۷۲	۲/۷۹۷	۳۴ درصد	۳۳ درصد
۹	U و RH, T _a	۲	۰/۸۴۵	۰/۸۵۶	۲/۷۱۸	۲/۷۳۴	۳۵ درصد	۳۳ درصد
۱۰	U و n, RH, T _a	۲	۰/۸۵۲	۰/۸۶۴	۲/۶۵۹	۲/۶۴۹	۳۵ درصد	۳۲ درصد
		۳	۰/۸۵۱	۰/۸۶۰	۲/۶۶۳	۲/۶۹۲	۳۳ درصد	۳۳ درصد

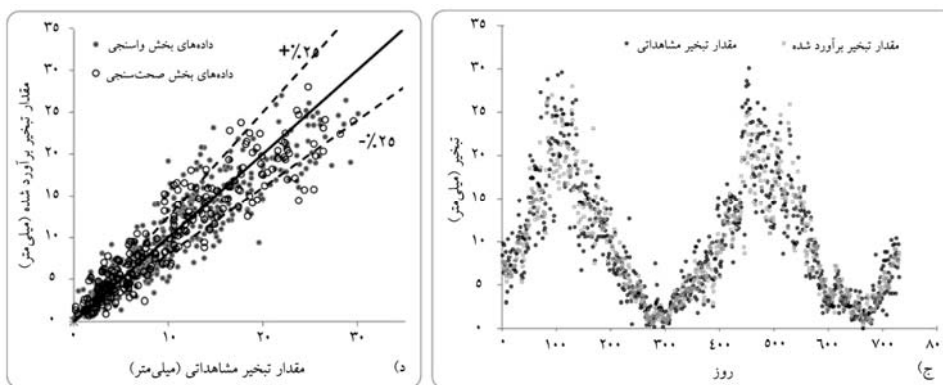
جدول ۷- ارزیابی مدل‌های ایجاد شده با استفاده از آماره‌های مختلف در ایستگاه ایزده.

ردیف	سناریوها	تعداد خوشه مناسب	R ^۲		RMSE		MAPE	
			واسنجی	صحت‌سنجی	واسنجی	صحت‌سنجی	واسنجی	صحت‌سنجی
۱	T _a	۲	۰/۸۷۲	۰/۸۸۵	۱/۸۴۴	۱/۶۸۱	۳۲ درصد	۲۸ درصد
۲	RH	۲	۰/۷۷۷	۰/۷۹۵	۲/۴۴۴	۲/۲۳۷	۳۷ درصد	۳۵ درصد
۳	n و T _a	۲	۰/۸۷۶	۰/۸۸۵	۱/۸۲۵	۱/۶۸۲	۳۰ درصد	۲۷ درصد
۴	U و T _a	۲	۰/۸۷۰	۰/۸۸۳	۱/۸۵۴	۱/۷۰۶	۳۲ درصد	۲۸ درصد
۵	n و RH	۳	۰/۸۶۴	۰/۸۷۶	۱/۹۰۰	۱/۷۵۴	۳۱ درصد	۲۹ درصد
۶	n و RH	۲	۰/۶۸۹	۰/۷۵۸	۳/۰۸۹	۲/۵۲۸	۴۱ درصد	۳۷ درصد
۷	U و RH	۲	۰/۷۷۸	۰/۷۹۸	۲/۴۳۹	۲/۲۲۶	۳۶ درصد	۳۶ درصد
۸	U و RH	۳	۰/۷۸۰	۰/۸۰۱	۲/۴۲۸	۲/۲۱۱	۳۶ درصد	۳۶ درصد
۹	RH و T _a	۲	۰/۸۹۱	۰/۹۱۵	۱/۶۹۷	۱/۴۴۴	۲۷ درصد	۲۴ درصد
۱۰	n و RH, T _a	۲	۰/۸۹۳	۰/۹۱۵	۱/۶۸۲	۱/۴۴۰	۲۷ درصد	۲۴ درصد
۱۱	n و RH, T _a	۳	۰/۸۹۵	۰/۹۱۹	۱/۶۶۷	۱/۴۱۶	۲۷ درصد	۲۴ درصد
۱۲	n و RH, T _a	۴	۰/۸۹۴	۰/۹۱۵	۱/۶۷۴	۱/۴۴۴	۲۷ درصد	۲۴ درصد
۱۳	U و RH, T _a	۲	۰/۸۸۸	۰/۹۱۱	۱/۷۲۲	۱/۴۷۲	۲۸ درصد	۲۵ درصد
۱۴	U و RH, T _a	۳	۰/۸۸۷	۰/۹۱۱	۱/۷۲۹	۱/۴۷۸	۲۸ درصد	۲۶ درصد
۱۵	U و n, RH, T _a	۲	۰/۸۹۲	۰/۹۱۴	۱/۶۹۲	۱/۴۵۱	۲۷ درصد	۲۴ درصد
۱۶	U و n, RH, T _a	۳	۰/۸۹۱	۰/۹۱۳	۱/۶۹۷	۱/۴۶۳	۲۷ درصد	۲۵ درصد

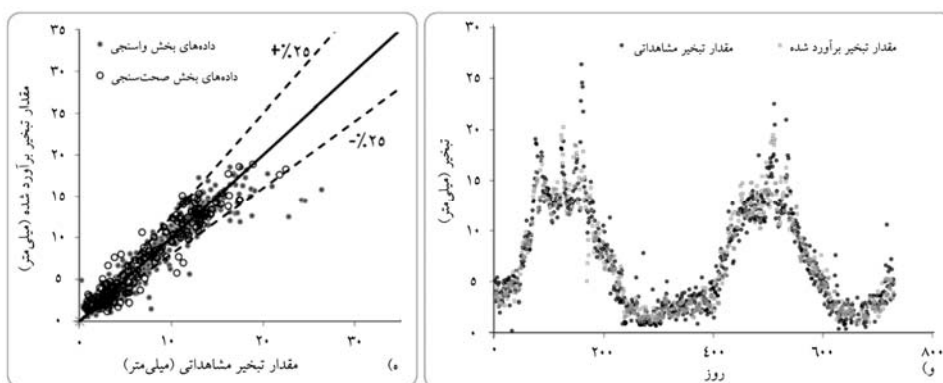
نتایج در ایستگاه اهواز نشان می‌دهد که بین دو سناریوی تک‌پارامتری موجود سناریوی شامل پارامتر ورودی دما با مربع ضریب همبستگی ۰/۷۸ در واسنجی و ۰/۷۴ در صحت‌سنجی نتیجه بهتری را به دست می‌دهد. همچنین بین سناریوهای دو پارامتری موجود سناریوهای ۷ و ۴ به ترتیب با مربع ضریب همبستگی ۰/۸۴ و ۰/۸۲ بهترین نتایج را ارائه می‌دهد. در ایستگاه اهواز بین سناریوهای بیش از دو پارامتر ورودی سناریوی ۱۰ با تعداد خوشه ۳ و مربع ضریب همبستگی حدود ۰/۸۶ بهترین نتیجه را می‌دهد (شکل ۳). در نهایت با توجه به نتایج به نظر می‌رسد ترتیب اهمیت نسبی پارامترهای ورودی در ایجاد بهترین مدل برای ایستگاه اهواز به صورت دمای هوا، رطوبت نسبی، سرعت باد و ساعات آفتابی است.



شکل ۳- نمودار نتایج به‌دست آمده از بهترین مدل (سناریوی ۱۰ و تعداد خوشه ۳) در ایستگاه اهواز.



شکل ۴- نمودار نتایج به‌دست آمده از بهترین مدل (سناریوی ۱۰ و تعداد خوشه ۲) در ایستگاه آبادان.



شکل ۵- نمودار نتایج به‌دست آمده از بهترین مدل (سناریوی ۸ و تعداد خوشه ۳) در ایستگاه ایذه.

نتایج در ایستگاه آبادان نیز نشان می‌دهد که سناریوی ۱۰ با بیش‌ترین تعداد پارامتر ورودی و با تعداد خوشه ۲ و مربع ضریب همبستگی حدود ۰/۸۶ بهترین نتیجه را داشته است (شکل ۴). البته میزان خطای نسبی در این ایستگاه بیش‌تر از اهواز است. با توجه به جدول ۶ به‌نظر می‌رسد ترتیب اهمیت نسبی پارامترهای ورودی در ایجاد بهترین مدل برای ایستگاه آبادان به‌صورت رطوبت نسبی، دمای هوا، سرعت باد و ساعات آفتابی است که می‌تواند ناشی از اقلیم مرطوب آن نسبت به اهواز باشد. ایستگاه ایذه نسبت به اهواز و آبادان در مدل‌سازی نتایج بهتری از خود نشان داده است (جدول ۷). در این ایستگاه سناریوی ۸ با تعداد خوشه ۳ و مربع ضریب همبستگی حدود ۰/۹۲ بهترین نتایج را به‌دست می‌دهد (شکل ۵). ترتیب اهمیت نسبی پارامترهای ورودی در ایجاد بهترین مدل در این ایستگاه مشابه ایستگاه اهواز است.

نتیجه‌گیری

با توجه به پیچیدگی و غیرخطی بودن پدیده تبخیر، نتایج به‌دست آمده از این پژوهش نشان می‌دهد که مدل استنتاج فازی تاکاجی - سوگنو با درک رفتارهای غیرخطی سیستم برای تخمین تبخیر از روی پارامترهای هواشناسی بسیار مناسب بوده و دارای دقت بالایی است. از طرفی از آنجایی‌که استخراج قوانین و پارامترهای بخش فرض و نتیجه در این مدل بسیار مشکل است، روش سیستماتیک خوشه‌بندی c-mean برای این منظور توانمندی قابل‌توجهی از خود نشان می‌دهد. نتایج نشان داد که به‌طورکلی در هر ۳ ایستگاه، در بیش‌تر سناریوها مجموعه داده‌ها به ۲ و ۳ خوشه تقسیم‌بندی شده است که این تعداد می‌تواند رابطه مستقیمی با خصوصیت اقلیمی منطقه داشته باشد. میزان خطای مدل‌سازی در ایستگاه آبادان بیش‌تر از اهواز است، در حالی‌که در ایستگاه ایذه نسبت به دو ایستگاه دیگر کم‌تر بوده و مدل، نتایج بهتری را در این ایستگاه از خود نشان می‌دهد. همچنین می‌توان نتیجه گرفت پارامترهای دما و رطوبت نسبی با توجه به اقلیم خوزستان از مهم‌ترین متغیرها در تخمین تبخیر از تشت می‌باشد. در مقام مقایسه، می‌توان به مطالعه رحیمی‌خوب (۲۰۰۶) اشاره کرد. پژوهش وی نیز نشان داد که مدل شبکه عصبی با پارامترهای ورودی شامل دمای هوا، رطوبت نسبی، سرعت باد و ساعات آفتابی، تبخیر از تشت را در استان خوزستان به‌خوبی تخمین می‌زند که نتایج آن تشابه زیادی با این پژوهش دارد. در نهایت پیشنهاد می‌شود از مدل‌های هوشمند دیگری مانند شبکه عصبی مصنوعی و یا ANFIS به‌منظور مقایسه با مدل ارائه شده در پیش‌بینی تبخیر از تشت استفاده گردد.

منابع

1. Bezdek, J. 1981. Pattern recognition with fuzzy objective function algorithms. Plenum. New York.
2. Keskin, M.E., Terzi, O., and Taylan, D. 2004. Fuzzy logic model approaches to daily pan evaporation estimation in western Turkey. *J. Hydr. Sci.* 49: 1001-1010.
3. Keskin, M.E., Terzi, O., and Taylan, D. 2009. Estimating daily pan evaporation using adaptive neural-based fuzzy inference system. *Theo. and Appl. Clim.* 98: 79-87.
4. Kim, D.W., Lee, K.H., and Lee, D. 2004. On cluster validity index for estimation of the optimal number of fuzzy cluster. *Patt. Recog. Lett.* 37: 2009-2025.
5. Koorehpazan-Dezfuli, A. 2008. Principles of fuzzy set theory and its applications in the modeling water engineering problems. Jahad Daneshgahi of the Amir-Kabir University. 261p. (In Persian)
6. Linacre, E.T. 1994. Estimating U.S. Class A Pan Evaporation from Few Climate Data. *Water Inter.* 19: 5-14.
7. Lohani, A.K., Goel, N.K., and Bhatia, K.K.S. 2005. Development of fuzzy logic based real time flood forecasting system for river Narmada in central India. International conference on innovation advances and implementation of flood forecasting technology.
8. Mahabir, C., Hicks, F.E., and Robinson Fayek, A. 2003. Application of fuzzy logic to forecast seasonal runoff. *Hydr. Proc.* 17: 3749-3762.
9. Moghaddamnia, A., Ghaffari Ghoushe, M., Piri, J., Amin, S., and Han, D. 2009. Evaporation estimation using artificial neural networks and adaptive neuro-fuzzy inference system techniques. *Adv. in Water Resour.* 32: 88-97.
10. Rahimi-Khoob, A. 2006. Estimating evaporation from limited climatic data using Artificial Neural Network modeling. 9th National Seminar on Irrigation and Evapotranspiration. Shahid-Bahonar Kerman University. (In Persian)
11. Tamer Ayvaz, M., Karahan, H., and Aral, M.M. 2007. Aquifer parameter and zone structure estimation using kernel-based fuzzy c-means clustering and genetic algorithm. *J. Hydrol.* 343: 240-253.
12. Vahidian-Kamiad, A., and Tareghian, H.R. 2002. Introduction of fuzzy logic for practical applications. Ferdowsi University. Press, 208p. (In Persian)
13. Wu, K.L., and Yang, M.S. 2005. A cluster validity index for fuzzy clustering. *Patt. Recog. Lett.* 26: 1275-1291.
14. Zhang, Y., Wang, W., Zhang, X., and Li, Y. 2008. A cluster validity index for fuzzy clustering. *Info. Sci.* 178: 1205-1218.



Gorgan University of Agricultural
Sciences and Natural Resources

J. of Water and Soil Conservation, Vol. 19(2), 2012
<http://jwfst.gau.ac.ir>

Application of fuzzy modeling based on clustering via c-mean for estimation of pan evaporation (case study: Khuzestan Province)

***S.R. Aghili¹, S. Boroomandnasab² and M. Kahe³**

¹M.Sc. Graduate, Dept. of Irrigation and Drainage, Shahid Chamran University of Ahvaz,

²Professor, Dept. of Irrigation and Drainage, Shahid Chamran University of Ahvaz,

³Ph.D. Student, Dept. of Water Structure, Shahid Chamran University of Ahvaz

Received: 2011/02/12; Accepted: 2011/11/01

Abstract

In this paper by using fuzzy logic and c-mean clustering, different combinations of daily climatic variables including air temperature, relative humidity, wind speed and sunshine hours as input data for predicting pan evaporation were investigated. Therefore, meteorological parameters in years of 1386 and 87 at three synoptic stations (Ahvaz, Abadan and Izeh) in Khuzestan province were used. The results showed that the Takagi-Sugeno fuzzy model and c-mean clustering were able to predict pan evaporation, so that fuzzy models with cluster number of 2 and 3 estimated pan evaporation parameter by correlation coefficients varying from 0.86 to 0.91.

Keywords: Pan evaporation, Takagi-sugeno fuzzy model, Fuzzy clustering, c-mean

* Corresponding Author; Email: aghili_roya@yahoo.com