



به نام خدا

دانشکده مهندسی برق و کامپیوتر، دانشکده فنی، دانشگاه تهران



پروژه نهایی درس شناسایی سیستم ها:

استفاده از خوشه های فازی با ساختار خطی در
شناسایی ساختار مدل های نورو فازی خطی -
محلی و کاربرد آن در پیش بینی میزان نیاز گیاه
به آب در شبکه های آبیاری

سید غلامحسین حاجی میرصادقی

شماره دانشجویی: 810187579

استاد درس: دکتر بابک نجار اعرابی

شهریور 1388

فهرست مطالب



چکیده	3
1 مقدمه	4
2 تخمین نیاز آبی گیاه	5
2.1 تبخیر-تعرق	5
2.2 روش های تخمین تبخیر-تعرق	5
2.2.1 روش های مستقیم تخمین تبخیر-تعرق	5
2.2.2 روش های غیر مستقیم تخمین تبخیر-تعرق	5
3 پیدا کردن خوشه های فازی با ساختار خطی	7
4 الگوریتم پیشنهادی	10
5 نتایج	12
5.1 شناسایی چند مورد معروف و مقایسه با مدل ها و روش های دیگر	12
5.1.1 تخمین دو تابع غیر خطی معروف	12
5.1.2 پیش بینی فعالیت های خورشیدی	16
5.2 پیش بینی میزان نیاز آبی گیاه در شبکه آبیاری سد حسنلو	17
5.2.1 پیش بینی سال های 2000 تا 2005	19
5.2.2 پیش بینی سال 1992	20
6 نتیجه گیری	23
مراجع	24



چکیده

در این پژوهش، روشی برای شناسایی مدل های نوروفازی خطی-محلی ارائه می شود. ایده ی کلی استفاده از الگوریتم های خوشه یابی فازی با ساختار خطی است که تقسیم بندی مناسبی را برای مدل های رگرسیون خطی-محلی به دست می دهند. برای این منظور، از روش خوشه بندی فازی c-varieties و نوع ترکیبی آن استفاده می شود. خوشه ها در فضایی تعمیم یافته که از ترکیب خروجی و ورودیها حاصل میشود شکل گرفته، سپس میانگین و واریانس وزندار (وزن ها همان درجات عضویت در خوشه ها هستند) هر خوشه در فضای ورودی محاسبه شده و به عنوان میانگین و واریانس کُرِنل های گوسی در مدل های خطی-محلی استفاده میشوند. در نهایت پارامتر های خطی تمام مدل ها با روش کمترین مربعات خطا محاسبه میشوند.

کارایی روش پیشنهادی برای شناسایی و تخمین توابع و سیستم های غیر خطی توسط چندین مثال شامل دو تابع غیر خطی معروف و همچنین سری زمانی لکه های خورشیدی که در تحلیل فعالیت های خورشیدی استفاده میشود مورد آزمون قرار گرفته و با الگوریتم ها و مدل های دیگر به ویژه الگوریتم LoLiMoT مقایسه می شود. در نهایت کاربرد این روش در پیش بینی کمیت تبخیر-تعرق که متناسب با میزان نیاز آبی گیاه می باشد و تخمین آن در جهت تخصیص دقیق آب در شبکه های آبیاری حائز اهمیت است بررسی می شود. نتایج نشان می دهد که روش ارائه شده در کاهش خطای شناسایی بسیار موثر است.



مدل های نورو-فازی خطی-محلی به دلیل توانایی بالای مدل سازی و سادگی تفسیر، یکی از بهترین مدل ها برای شناسایی سیستم های غیرخطی هستند [1]. اما سخت ترین مرحله در استفاده از این سیستم ها شناسایی ساختار مناسب برای آنها شامل محدوده و شکل مدل های خطی محلی و توابع اعتبار¹ آنها می باشد. ایده استفاده از خوشه یابی برای تشکیل مدل های خطی محلی برای تقسیم بندی هوشمندانه تر و کاهش تعداد مدل ها، ایده ی جدیدی نیست. مثلاً در [2] از روش خوشه یابی Gustafson-Kessel برای شناسایی مدل های تاکاگی-سوگنو² استفاده شده است. در [3] نیز روش اصلاح شده ی GATHA-GEVA مورد استفاده قرار گرفته که در آن پارامتر های توابع اعتبار مستقیماً از پارامتر های خوشه یابی به دست می آیند. نمونه های دیگری از کاربرد خوشه یابی برای شناسایی ساختار را می توان در [4] و [5] یافت. علاوه بر روش های خوشه یابی، الگوریتم های دیگری مانند LoLiMoT [6] نیز به منظور شناسایی مدل های خطی-محلی به کار می روند. در [7] روش جدیدی مانند LoLiMoT ارائه شده است که امکان تقسیم و ترکیب مدل ها در آن لحاظ شده است و بنابراین نیازی به تعریف خانواده مدل توابع اعتبار یا شکل خوشه ها ندارد. به علاوه این روش و همینطور LoLiMoT بر خلاف روش های خوشه یابی که عمدتاً بر مبنای گرادیان بوده و رفتار محلی دارند، رویکردی غیر گرادیانی و ساده داشته و نیاز به تکرار فرآیند های بهینه سازی ندارند (البته این الگوریتم ها نیز محدودیت های خودشان را دارند).

ما در این تحقیق به دنبال الگوریتم های خوشه یابی ای هستیم که تا حد امکان روابط خطی بین ورودی و خروجی را برای ما کشف کنند. به همین دلیل، ما از خوشه یابی فازی c-varieties و نوع ترکیبی آن که برای تشخیص زیرساختارهای خطی در داده ها پیشنهاد شده اند استفاده کرده و روشی برای شناسایی پارامتر های توابع اعتبار و مدل های خطی-محلی ارائه می کنیم. در پایان، الگوریتم پیشنهادی را برای پیش بینی سری زمانی کمیت تبخیر-تعرق³ در شبکه آبیاری شد حسنلو به کار می بریم. تبخیر-تعرق از مهمترین معیارهای تشخیص میزان نیاز آبی گیاه بوده و در دهه ی اخیر، استفاده از شبکه های عصبی برای پیش بینی آن بسیار رواج یافته است [8]-[11].

در ادامه ی این گزارش، مباحث به این ترتیب ارائه می شوند. ابتدا در بخش 2 توضیحاتی درباره ی نیاز آبی گیاه، کمیت تبخیر-تعرق و روش های تخمین آن ارائه میشود. سپس در بخش 3 مختصری درباره روش های خوشه یابی با ساختار های خطی مانند روش های فازی c-lines، c-varieties و مدل ترکیبی آنها صحبت می شود. بخش 4 به شرح الگوریتم پیشنهادی برای شناسایی ساختار مدل های نوروفازی خطی-محلی می پردازد. نتایج شبیه سازی بر روی چند مثال معروف و همینطور پیش بینی سری زمانی کمیت تبخیر-تعرق به همراه مقایسه با روش های دیگر در بخش 5 ارائه می شوند. نهایتاً بخش 6 به جمع بندی و نتیجه گیری نتایج ارائه شده اختصاص یافته است.



¹ Validity function

² Takagi-Sugeno

³

2 تخمین نیاز آبی گیاه

یکی از اساسی ترین مراحل برنامه ریزی آبیاری در شبکه های آبیاری و زهکشی، برآورد نیاز مصرف کنندگان آب (بر اساس الگوی کشت یا الگوی نیاز) می باشد. از آنجا که تبخیر-تعرق رابطه مستقیم با نیاز آبی گیاه دارد، برآورد تبخیر-تعرق اصلیتین گام در جهت تخصیص دقیق آب در شبکه ها می باشد.

2.1 تبخیر-تعرق

در زنجیره ی آب-خاک-گیاه-اتمسفر، آب مستقیماً از سطح خاک و یا توسط گیاه به داخل اتمسفر وارد می شود انتقال آب از سطح خاک به هوا را تبخیر و خارج شدن آن از گیاه را تعرق گویند. این دو پدیده هر دو ماهیت تبخیری داشته و چون تفکیک آنها از یکدیگر امکان پذیر نمیباشد، مجموعاً به نام تبخیر-تعرق در نظر گرفته شده و با علامت ET نشان داده می شود.

تبخیر-تعرق پتانسیل حداکثر مقدار آبی است که اگر محدودیتی وجود نداشته باشد، می تواند توسط سطوح خاک و گیاه از خاک و گیاه وارد جو شود. تبخیر-تعرق گیاه مرجع همان تبخیر-تعرق پتانسیل برای یک پوشش گیاهی خاص است که معمولاً چمن یا یونجه انتخاب می شود. تعریفی که برای گیاه مرجع چمن شده است، این است که ارتفاع این گیاه 8 تا 15 سانتی متر بوده، سطح وسیعی را در بر گرفته و بطور کامل و یکنواخت زمین را پوشش داده باشد، سبز و شاداب بوده و بدون محدودیت آب، تبخیر-تعرق آن صورت گیرد.

2.2 روش های تخمین تبخیر-تعرق

روشهایی که برای تخمین تبخیر-تعرق بکار برده می شود در دو گروه اصلی قرار می گیرند که عبارتند از روشهای مستقیم و روشهای محاسبه ای. در روشهای مستقیم بخش کوچک و کنترل شده ای از مزرعه را مجزا کرده و مقدار تبخیر-تعرق در یک دوره زمانی مستقیماً اندازه گیری می شود. حال آنکه در روشهای محاسبه ای که می توان آنها را روشهای غیرمستقیم دانست از عوامل مختلف اقلیمی و گیاهی استفاده شده و از روی ارتباط آنها با تبخیر-تعرق و معادلهایی که قبلاً با روشهای مستقیم واسنجی شده اند، تبخیر-تعرق پوشش گیاهی مورد نظر تخمین زده می شود. البته می بایست به این نکته اشاره کرد که هیچکدام از این روشها نمی توانند تبخیر-تعرق را بطور دقیق برآورد نمایند ولی برخی از آنها در بعضی مناطق نتایجی را بدست میدهند که بیشتر با واقعیت مطابقت دارد. از نظر عملی روشی مطلوب است که اولاً آسان بوده و ثانیاً نتایج حاصل از آن واقعیتر باشد.

2.2.1 روش های مستقیم تخمین تبخیر-تعرق

معمولترین روش مستقیم تعیین تبخیر-تعرق، استفاده از اصل بیلان جرمی در یک حجم کنترل شده از خاک است. در این روش رطوبت خاک در ابتدا و انتهای دوره معینی اندازه گیری و مقادیر آب ورودی و خروجی از زمین اندازه گیری و از روی آنها تبخیر-تعرق اندازه گیری میشود. در مقیاس کوچک ساده ترین وسیله جهت اندازه گیری تبخیر-تعرق، لایسیمتر است.

2.2.2 روش های غیر مستقیم تخمین تبخیر-تعرق

روشهای غیر مستقیمی که برای محاسبه ET0 (تبخیر-تعرق پتانسیل) پیشنهاد شده است را می توان بطور کلی در 4 گروه تقسیم کرد که عبارتند از:

(1) روشهای موسوم به آیرودینامیک

(2) روشهای موسوم به توازن انرژی



3) روشهایی که از ترکیب دو روش فوق حاصل شده و روشهای ترکیبی خوانده می شوند.

4) روشهای تجربی

در روشهای آبرودینامیک میزان تبخیر-تعرق تابعی از سرعت باد و اختلاف فشار بخار بین لایه هوای چسبیده به برگ و هوای اطراف دارد. این معادلات فقط در شرایط تحقیقاتی قابل استفاده می باشد زیرا اولاً اندازه گیری سرعت باد و دانسیته بخار آب می بایست با دقت صورت گیرد که کاری بسیار مشکل است ثانیاً نیاز به دستگاههایی دارد که هزینه تهیه آنها زیاد است و ثالثاً این دستگاه ها نیازمند افراد کار آزموده است.

در روشهای توازن انرژی، میزان تبخیر-تعرق از یک سطح مرطوب گیاه و خاک، توسط انرژی موجود در سطح کنترل میشود. انرژی موجود در سطح تابعی از تابش ورودی به سطح مورد نظر و گرمایی است که به صورت جابجایی افقی از اطراف به سطح مورد نظر وارد می شود. در روش توازن انرژی نیز به دلیل اینکه نمی توان در آن نسبت بوون را به دقت برآورد کرد، همانند روشهای آبرودینامیک جنبه تحقیقاتی داشته و در عمل استفاده چندانی از آن نمی شود.

در روشهای تجربی رابطه بین ETO و یک یا چند عامل هواشناسی اساسی مؤثر بر تبخیر-تعرق در نظر گرفته شده است. دمای هوا، تابش خورشید، رطوبت هوا و سرعت باد از جمله این متغیرها به شمار میروند. این روشها گرچه ساده می باشند اما کاربرد آنها برای دوره های کوتاهتر از ماهانه توصیه نمیشود.

در سال 1984 پنمن دانشمند انگلیسی از ترکیب روشهای آبرودینامیک و توازن انرژی روشی را برای محاسبه تبخیر-تعرق ارائه نمود که به روش ترکیبی یا معادله پنمن معروف شد. معادله پنمن در ابتدا برای محاسبه تبخیر از سطح آزاد آب ارائه گردید و سپس برای محاسبه شدت تبخیر-تعرق از یک سطح کامل پوشیده از چمن مورد استفاده قرار گرفت.

معادله ی پنمن-فائو که توسط متخصصان فائو ارائه گردید هنوز هم به عنوان یکی از معادلات کاربردی در محاسبات تبخیر-تعرق گیاه مرجع مورد استفاده قرار می گیرد. ولی از آنجا که در این معادله فرض شده است که تبخیر-تعرق فقط توسط عوامل آب و هوایی کنترل شده و نقش خود گیاه در آن در نظر گرفته نشده است لذا به تدریج اهمیت خود را از دست داده و معادلات دیگری جای آن را گرفته اند.

روش فائو-پنمن-مانتیت شناخته شدهترین و معتبرترین این روشها می باشد که در حال حاضر کاربرد گسترده ای در مناطق مختلف آب و هوایی دارد. در این معادله فاکتورهای مقاومت سطح (ra و rc که به ترتیب مقاومت پوشش گیاهی و مقاومت آبرودینامیک لایه هوای مجاور پوشش گیاهی نام دارند) اضافه شده اند. معادله فائو پنمن مانتیت به عنوان روش استاندارد محاسبه نیاز آبی گیاهان شناخته شده است. این روش هم در شرایط خشک و هم در شرایط مرطوب نتایج قابل قبولی را ارائه می دهد.



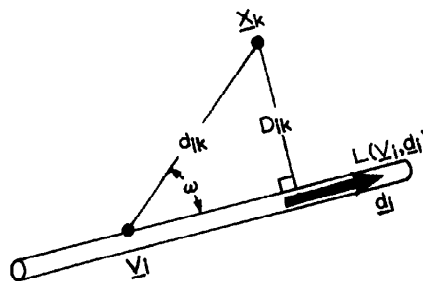
3 پیدا کردن خوشه های فازی با ساختار خطی

معروف ترین روش تشخیص بخش های فازی مقید، روش خوشه یابی FCM می باشد که از میانگین خوشه ها به عنوان نشان دسته آنها استفاده میکند. این روش، داده ها را براساس نزدیکی در فضای اقلیدسی خوشه بندی نموده و برای داده های با ساختار کروی مناسب است. همانطور که قبلا ذکر شد، ما در این مطالعه به دنبال روش خوشه یابی موثری برای شناسایی مدل های نوروفازی خطی-محلی هستیم. بنابراین اینگونه به نظر می رسد که روش های خوشه یابی که قادر به تشخیص روابط و ساختار های خطی در داده ها هستند میتوانند بسیار کارآمد باشند. ما در ادامه روش های فازی c-lines (FCL)، فازی c-varieties (FCV) و یک مدل ترکیبی را که در [12] و [13] پیشنهاد شده اند به طور مختصر توضیح داده و در بخش بعدی نحوه ی استفاده از آنها را برای شناسایی ساختار در مدل های نوروفازی خطی-محلی بیان میکنیم.

در سال 1981 دو مقاله کلیدی توسط بزدرک⁴ و همکاران در مورد تشخیص و شناسایی زیرساختار های خطی در داده ها ارائه شد [12]، [13]. مقاله اول به توضیح الگوریتم FCL اختصاص داشته و هدف آن پیدا کردن تعدادی خط برای پوشش دادن داده ها در فضای تعریفی آنها است. در واقع نشان دسته ی هر خوشه، یک خط می باشد که با یک نقطه روی آن (v) و بردار جهت آن (v) مشخص می شود. مثلاً در فضای S بُعدی داریم

$$L(v; d) = \{y \in R^S | y = v + td; t \in R\}$$

که در شکل زیر نمونه ای از این خط ها دیده می شود:



شکل 1- نمونه ای از باز نمایی خط در روش FCL

الگوریتم کلی تری که در مقاله ی دوم، بزدرک به آن پرداخته است، الگوریتم FCV است که تعدادی فوق صفحه با بُعد یکسان برای پوشش دادن فضای داده ها پیدا مینماید. در این روش، نشان دسته ی هر خوشه یک فوق صفحه با بعد r ($0 \leq r < s$) در فضای R^S است که با یک نقطه روی آن و بردار که آنرا span میکنند معرفی میشود:

$$V_r = (V_{r1}, V_{r2}, \dots, V_{rc})$$

$$V_{ri}(v_i; d_{i1}, d_{i2}, \dots, d_{ir}) = \left\{ y_i \in R^S \mid y_i = v_i + \sum_{j=1}^r t_{ji} d_{ji}; t_{ji} \in R \right\}$$

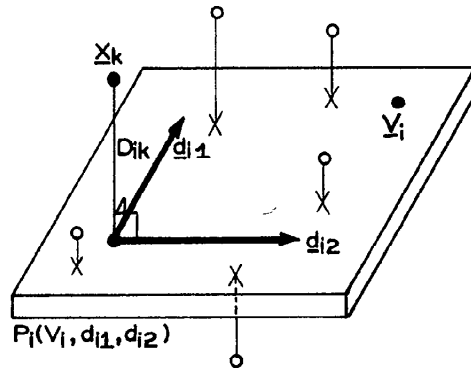
که c برابر تعداد خوشه ها و v_i یک نقطه بر روی فوق صفحه i ام و $d_{i1} \dots d_{ir}$ مجموعه بردارهایی هستند که آن فوق صفحه را span می کنند. برای پیدا کردن این مجموعه نقاط و بردار ها در قالب یک مسئله خوشه بندی فازی مقید، از بهینه سازی تابع هزینه ی زیر استفاده می شود.

$$J_{Vrm}(U, V) = \sum_{k=1}^n \sum_{i=1}^c (u_{ik})^m (D_{ik})^2$$

که در آن u_{ik} مشخص کننده درجه عضویت داده k ام در خوشه i ام بوده و D_{ik} فاصله آن داده را از آن خوشه نشان می دهد که با فرمول زیر محاسبه می شود:

$$D_A(\mathbf{x}_k, \mathbf{V}_{ri}) = \left(\|\mathbf{x}_k - \mathbf{v}_i\|_A^2 - \sum_{j=1}^r ((\mathbf{x}_k - \mathbf{v}_i, \mathbf{d}_{ij})_A)^2 \right)^{1/2} = D_{ik}$$

که A یک ماتریس معین مثبت متقارن دلخواه می باشد و $\langle x, y \rangle_A = x^T A y$ و $\|x\|_A^2 = x^T A x$ برای حالت خاص فضای 3 بعدی، نمونه ای از صفحات به دست آمده با الگوریتم FCV در شکل زیر به نمایش در آمده است.



شکل 2- نمونه ای از بازنمایی صفحه در روش FCV

از آنجا که الگوریتم FCV روش کلی تری نسبت به FCL می باشد، ما در این نوشتار از بحث پیرامون FCL اجتناب کرده و فقط به شرح FCV می پردازیم. نوع دیگری از خوشه یابی که در مقاله ی دوم بزدک مطرح شده است، نوع ترکیبی FCV است که در آن مجموعه ای از مجموعه FCV های با بُعد مختلف در نظر گرفته میشود. بدین معنی که در این خوشه بندی سعی می شود که داشتن روابط خطی در چند بُعد مختلف به عنوان معیار نزدیکی داده ها مورد استفاده قرار گیرد و به همین دلیل، تابع هزینه به صورت مجموع وزننداری از تابع هزینه های هر بعد تعریف می شود.

فرض کنید که $\{r_i\}$ مجموعه ای از p عدد صحیح متفاوت، $\{\alpha_i\}$ مجموعه ای از p عدد در بازه 0 و 1 که مجموع آنها برابر 1 است باشند، و همینطور $\{\mathbf{V}_{ri}\}$ مجموعه ای از مجموعه های c-varieties به شکل

$$\mathbf{V}_{ri} = (V_{ri1}, V_{ri2}, \dots, V_{ric}) \quad \dim(V_{ri1}) = r_i$$

باشد، آنگاه تابع بهینه سازی ترکیبی به صورت زیر تعریف می گردد.

$$J_{\{\mathbf{V}_{ri}\}m}^{\{\alpha_i\}}(U, \mathbf{V}_{r1}, \dots, \mathbf{V}_{rp}) = \sum_{i=1}^p \alpha_i J_{\mathbf{V}_{ri}m}$$

حالت خاصی از نوع ترکیبی که مورد استفاده ی ما در این پژوهش میباشد، حالت دو مجموعه ای آن به صورت J_{pqm}^α و به خصوص $J_{V_{0q}m}^\alpha$ می باشد. یعنی ما می خواهیم علاوه بر داشتن خوشه هایی با روابط خطی، داده های هر خوشه از لحاظ اقلیدسی نیز به هم نزدیک باشند تا در هنگام تخمین مدل های خطی-محلی، استفاده از کرنل های گوسی که رفتار آنها بر مبنای فاصله ی اقلیدسی از مرکز آنها است، معقول باشد. تابع بهینه سازی برای این مسئله به صورت زیر تعریف می شود.

$$J_{V_{0q}m}^\alpha = (1 - \alpha) J_{V_{0m}} + \alpha J_{V_{qm}}$$



که J_{vom} همان J_{mm} می باشد که به عنوان تابع هزینه در الگوریتم FCM استاندارد استفاده میشود.

الگوریتم های FCL، FCV و ترکیبی (که برای حالت خاص توضیح داده شده از این پس با نماد FCV_{0q} که q معرف بُعدِ ترمِ دوم (بخش خطی) در تابع هزینه می باشد نمایش داده میشود) با جزییات کامل شامل حالت های ویژه، قضایا و اثبات، تحلیل سرعت همگرایی و مثال های متنوع در مقالات [12] و [13] به تفصیل شرح داده شده اند.



4 الگوریتم پیشنهادی

ایده کلی به این صورت است که ابتدا ورودی که بُعدی برابر S دارد با خروجی یک بُعدی ترکیب شده و یک فضای تعمیم یافته ی $1 + S$ بُعدی می سازند. سپس خوشه یابی در این فضا با $Jvsm$ (یا $Jvsm^\alpha$) انجام می شود. به این ترتیب رابطه خطی بین ورودی و خروجی لحاظ شده (به اضافه نزدیکی اقلیدسی داده ها برای نوع ترکیبی)، و در ادامه می توان میانگین و واریانس کرنل های گوسی (توابع اعتبار فازی) در مدل های خطی-محلی را از میانگین و واریانس وزندار خوشه ها (متناسب با درجه عضویت آنها) در فضای ورودی تخمین زد. مشروح الگوریتم پیشنهادی به قرار زیر است

(1) شروع: از یک مدل خطی تک شروع می کنیم که همان تخمین مربعات خطای خطی از فضای ورودی به خروجی می باشد. سپس برای هر نمونه داده، بردار ورودی x_k را که S بُعدی است با خروجی متناظر آن y_k ترکیب نموده و یک بردار تعمیم یافته به صورت $z_k = [x_k, y_k]$ می سازیم.

(2) تعداد مدل های خطی-محلی را یکی اضافه کرده و اعمال زیر را برای هر مجموعه از این مدل ها انجام می دهیم

a. مجموعه ی حاصل از بردار های تعمیم یافته را با الگوریتم خوشه یابی توضیح داده شده در بخش قبل به تعداد مدل های خطی-محلی این مرحله تقسیم بندی می کنیم و به این ترتیب u_{ik} ، v_i و d_{ij} را برای هر داده k ، خوشه i و بُعد j به دست می آوریم

b. برای هر خوشه، میانگین و واریانس را در فضای برداری ورودی (و نه فضای تعمیم یافته) به صورت زیر محاسبه می کنیم و آن ها را به عنوان میانگین و واریانس کرنل های گوسی مربوط به هر LLM در نظر می گیریم

$$\hat{c}_i = \frac{\sum_{k=1}^n (\hat{u}_{ik})^m x_k}{\sum_{k=1}^n (\hat{u}_{ik})^m}$$

$$\hat{\sigma}_{ir}^2 = \frac{\sum_{k=1}^n (\hat{u}_{ik})^m (x_{kr} - \hat{c}_{ir})^2}{\sum_{k=1}^n (\hat{u}_{ik})^m}$$

که n معرف تعداد کل نمونه هاست، \hat{c}_i بردار S بُعدی و نشان دهنده ی میانگین تخمین زده شده برای مدل خطی-محلی i ام است، و $\hat{\sigma}_{ir}^2$ واریانس تخمین زده شده برای بُعد r ام ($r = 1 \dots S$) مدل خطی-محلی i ام است.

c. پارامتر های مجموعه مدل های خطی را به صورت جهانی و به کمک تخمین مربعات خطای خطی با شرایط و روابط زیر شناسایی می کنیم.

اگر پارامتر های یادگیری برای مدل LLNF به قرار زیر بوده

$$\omega = [\omega_{10} \ \omega_{11} \ \dots \ \omega_{1s} \ \omega_{20} \ \omega_{21} \ \dots \ \omega_{M0} \ \dots \ \omega_{Ms}]^T$$

که M برابر تعداد مدل هاست و مدل به صورت زیر نمایش داده شود

$$\hat{y} = X\hat{\omega}$$

که در آن

$$X = [X_1 \ X_2 \ X_3 \ \dots \ X_M]$$



$$X_i = \begin{bmatrix} \phi_i(u(1)) & u_1(1)\phi_i(u(1)) & \dots & u_s(1)\phi_i(u(1)) \\ \phi_i(u(2)) & u_1(2)\phi_i(u(2)) & \dots & u_s(2)\phi_i(u(2)) \\ \vdots & \vdots & \ddots & \vdots \\ \phi_i(u(n)) & u_1(n)\phi_i(u(n)) & \dots & u_s(n)\phi_i(u(n)) \end{bmatrix}$$

9

$$u = [u_1 \ u_2 \ \dots \ u_s]^T$$

آنگاه پارامترهای شناسایی به صورت زیر محاسبه می شوند

$$\hat{\omega} = (X^T X)^{-1} X^T y$$

d. خطای تخمین را بر روی داده های یادگیری و تست محاسبه کرده و ذخیره می کنیم، و سپس مرحله 2

را تا رسیدن به حداکثر تعداد مدل های خطی-محلی تکرار می کنیم.

(3) نمودار خطای یادگیری و تست را در یک نمودار رسم کرده و با توجه به روند تغییرات مدل بهینه را انتخاب می کنیم. این مدل زمانبست که روند خطای تست شروع به افزایش کند یا برای مدت طولانی ثابت بماند.

از آنجا که الگوریتم خوشه یابی به کار رفته مبتنی بر گرادیان مزدوج بوده و محلی می باشد، با یک بار انجام آن نمیتوان به جواب های بهینه ی جهانی دست یافت (اگرچه که نتایج شبیه سازی ما نشان میدهد می توان به جواب های قابل قبولی رسید). برای رفع این مشکل در مرحله 2 از الگوریتم بالا خوشه یابی را چند بار با شرایط اولیه ی متفاوت تکرار کرده و بهترین خوشه بندی که منتج به کمترین خطای تست می شود را به عنوان تقسیم بندی بهینه برای آن تعداد LLM در نظر میگیریم.

نکته ی آخر اینکه برای انتخاب بین الگوریتم FCV و FCV ترکیبی برای خوشه یابی ($FCV_{0.5}$) و همینطور مقادیر α در الگوریتم ترکیبی از روش های متداول تست اعتبار (validation) استفاده می کنیم.



5 نتایج

در این بخش ما توانایی الگوریتم پیشنهادی را برای تخمین و شناسایی چندین تابع و سیستم غیر خطی مطالعه می کنیم. ابتدا از دو تابع غیر خطی معروف و یک سری زمانی که برای تست مدل ها و الگوریتم های شناسایی و تخمین به کار گرفته میشوند استفاده می کنیم و سپس به پیش بینی سری های زمانی تبخیر-تعرق که در تخمین نیاز آبی گیاه مورد مطالعه قرار می گیرند می پردازیم.

5.1 شناسایی چند مورد معروف و مقایسه با مدل ها و روش های دیگر

5.1.1 تخمین دو تابع غیر خطی معروف

تابع اول، یک تابع غیر خطی با دو ورودی و یک خروجی است که به سینک دو بعدی معروف بوده و در [14] مورد مطالعه قرار گرفته است. این تابع به صورت زیر نمایش داده شده

$$z = \text{sinc}(x, y) = \frac{\sin(x)}{x} \times \frac{\sin(y)}{y} \quad (5.1)$$

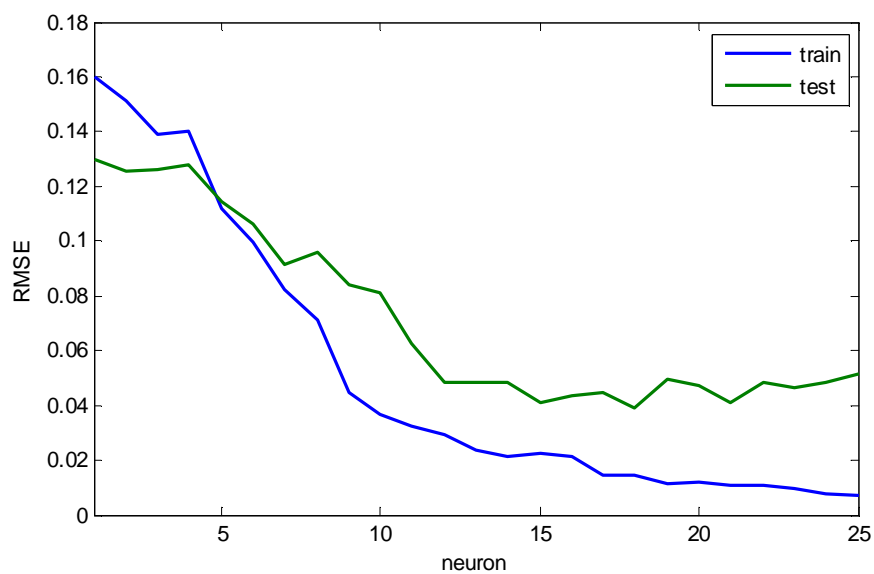
و در فضای $[-10, 10] \times [-10, 10]$ تعریف می شود. برای مرحله ی یادگیری 121 جفت ورودی و خروجی به صورت تصادفی استخراج شده و برای مرحله تست نیز به همین ترتیب 71 جفت ورودی و خروجی انتخاب می شوند. نتایج شبیه سازی با استفاده از الگوریتم پیشنهادی و الگوریتم LoLiMoT برای یادگیری مدل های نورو فازی خطی-محلی و همینطور مدل های دیگری همچون MLP و RBF برای تخمین این تابع در جدول 1 آورده شده است. در شبیه سازی الگوریتم پیشنهادی از FCV ترکیبی با $\alpha = 0.7$ استفاده شده و برای هر تعداد LLM عمل خوشه یابی 20 بار تکرار شده و بهترین نتیجه در نظر گرفته شده است.

جدول 1- نتایج تخمین تابع غیر خطی (5.1) با روش های مختلف

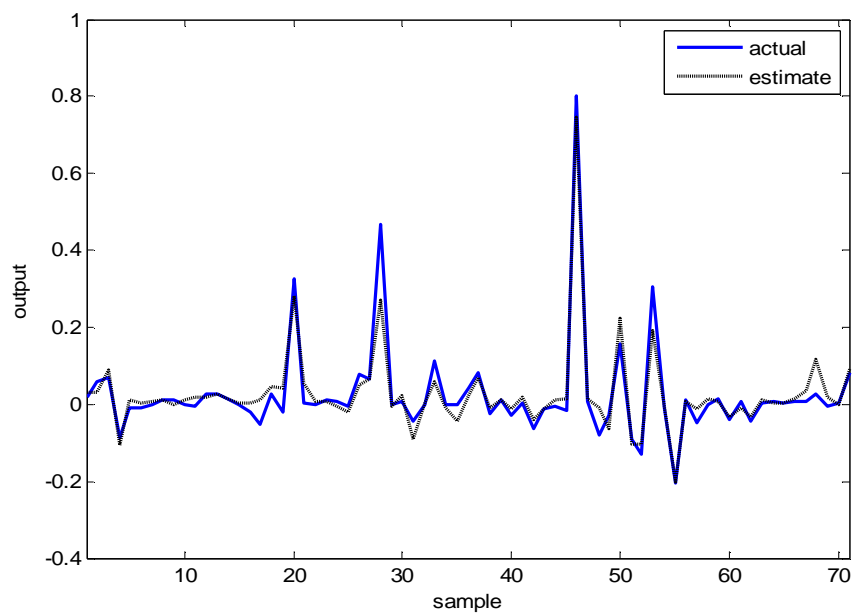
الگوریتم یا مدل	FCV	LoLiMoT	MLP	RBF
مجذور میانگین مربع خطای یادگیری	0.0071	0.0441	0.0069	0.1109
مجذور میانگین مربع خطای تست	0.0388	0.0661	0.0529	0.1145
تعداد نورون های میانی یا LLM	18	12	25	17

همانطور که در جدول 1 دیده می شود خطای تست برای الگوریتم پیشنهادی از دیگران کمتر بوده و مدل بهتری از تابع را به دست می دهد. نمودار تغییرات خطای یادگیری و تست به ازای تعداد نورون های مختلف با الگوریتم خوشه یابی FCV ترکیبی در شکل 3 نمایش داده شده است. به علاوه، برای بهترین مدل به دست آمده، نمودار خروجی واقعی و خروجی تخمین زده شده و همینطور نمودار خطای تخمین بر روی داده های تست در شکل های 4 و 5 نشان داده شده اند.



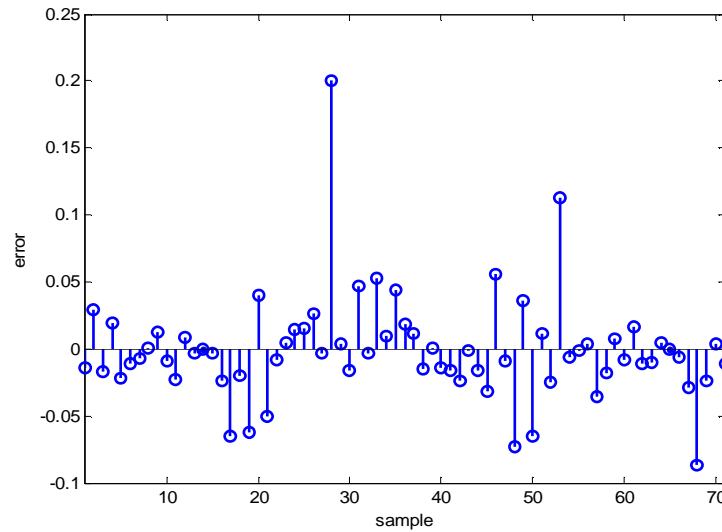


شکل 3- خطای تخمین در الگوریتم پیشنهادی به ازای تعداد نورون (مدل های خطی-محلی) برای داده های یادگیری و تست



شکل 4- نمودار خروجی واقعی و خروجی تخمین زده شده بر روی داده ی تست





شکل 5- نمودار خطای تخمین بر روی داده تست

تابع دوم، تابع غیر خطی معروفی است که در [14]، [15] و [16] برای بررسی و تایید کارآمدی الگوریتم ها و مدل های شناسایی مطرح شده به کار رفته است. این تابع که دارای 3 ورودی است بدین صورت تعریف می شود:

$$y = (1 + x_1^{0.5} + x_2^{-1} + x_3^{-1.5})^2 \quad (5.2)$$

برای انتخاب داده های یادگیری و تست از بازه برداری $[1, 6] \times [1, 6] \times [1, 6]$ ، داده ها به صورت تصادفی استخراج می شوند. در این مسئله 216 نمونه برای یادگیری و 125 نمونه برای تست در نظر گرفته میشوند و در فرایند شبیه سازی الگوریتم ها مورد استفاده قرار می گیرند.

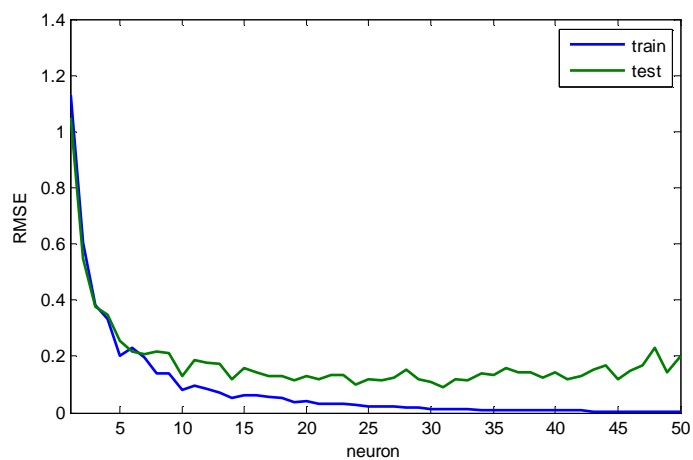
برای این تابع، ما فقط مدل های نورو فازی خطی-محلی را در نظر گرفته و الگوریتم پیشنهادی را با LoLiMoT مقایسه می کنیم. نتایج این مقایسه در جدول 2 ارائه شده اند. همانطور که دیده می شود کارایی الگوریتم پیشنهادی در این مسئله بهتر از الگوریتم LoLiMoT می باشد. توضیح اینکه در شبیه سازی الگوریتم پیشنهادی از خوشه یابی FCV استفاده شده و برای هر تعداد LLM عمل خوشه یابی 30 بار تکرار شده و بهترین نتیجه در نظر گرفته شده است.

جدول 2- نتایج تخمین تابع غیر خطی (5.2) با روش های مختلف

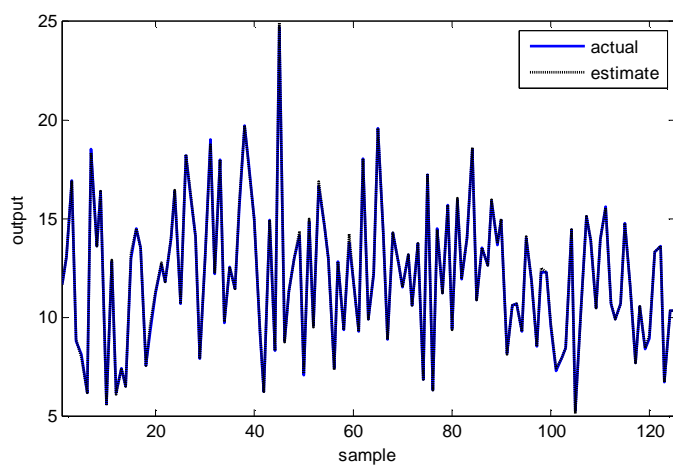
الگوریتم	FCV	LoLiMoT
مجذور میانگین مربع خطای یادگیری	0.0018	0.0742
مجذور میانگین مربع خطای تست	0.0913	0.1407
تعداد مدل های خطی-محلی (LLM)	31	30

نمودار تغییرات خطای یادگیری و تست به ازای تعداد نورو ن های مختلف با الگوریتم خوشه یابی FCV در شکل 6 نمایش داده شده است. به علاوه، برای بهترین مدل به دست آمده، نمودار خروجی واقعی و خروجی تخمین زده شده و

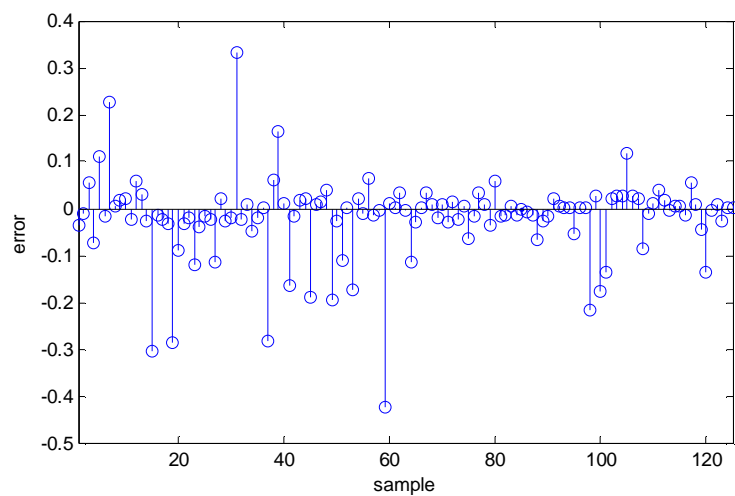




شکل 6- خطای تخمین در الگوریتم پیشنهادی به ازای تعداد نورون (مدل های خطی-محلی) برای داده های یادگیری و تست



شکل 7- نمودار خروجی واقعی و خروجی تخمین زده شده بر روی داده ی تست



شکل 8- نمودار خطای تخمین بر روی داده تست



5.1.2 پیش بینی فعالیت های خورشیدی

ما در این قسمت به پیش بینی یکی از سری های زمانی معروف در زمینه فعالیت های خورشیدی میپردازیم. تعداد لکه های خورشیدی یکی از معیار های خوب برای اندازه گیری فعالیت خورشیدی میباشد که از فرمول Wolf محاسبه می شود:

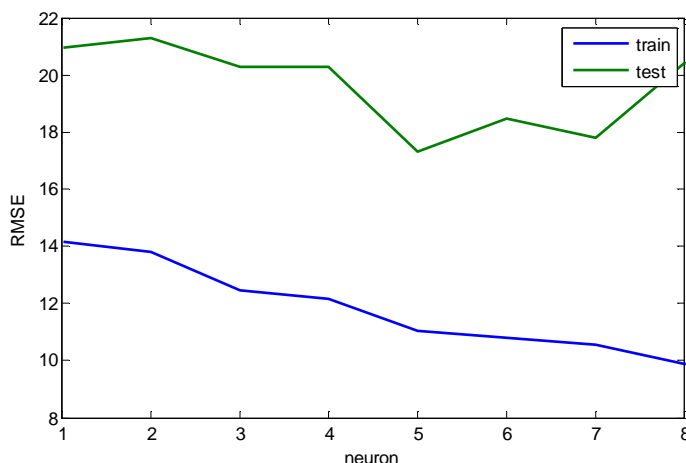
$$R = k(10g + s) \quad (5.3)$$

که g معرف تعداد گروه های لکه های خورشیدی بوده و s تعداد کل لکه ها در تمام گروه ها را نشان میدهد و k یک فاکتور برای scaling می باشد که به شرایط مشاهده مربوط است. ما در اینجا از داده های سالانه استفاده کرده و نتایج را با نتایج به دست آمده در مقاله ی رضایی یوسفی و همکاران [17] که از LoLiMoT استفاده کرده است مقایسه می کنیم. تعداد 231 نمونه از سال 1700 تا 1930 به عنوان داده های یادگیری انتخاب شده و بقیه تا سال 2005 برای تست در نظر گرفته شده اند. برای انتخاب ورودی های مناسب از نتایج مقاله [17] استفاده می کنیم. در این مقاله با استفاده از روشی مبتنی بر mutual information از میان lag 15 اول lag 5 که مهمترین تاثیر را در خروجی دارند انتخاب می شوند. این lag ها به ترتیب از لحظات t-1، t-5، t-10، t-11، t-2 استخراج می شوند. نتایج شبیه سازی با استفاده از الگوریتم پیشنهادی و مقایسه ی آن با الگوریتم LoLiMoT از [17] در جدول 3 نمایش داده شده است. برای این مسئله از خوشه یابی FCV ترکیبی با $\alpha = 0.8$ استفاده شده و برای هر مدل 20 بار عمل خوشه یابی تکرار گردیده است.

جدول 3- نتایج پیش بینی سری زمانی لکه های خورشیدی

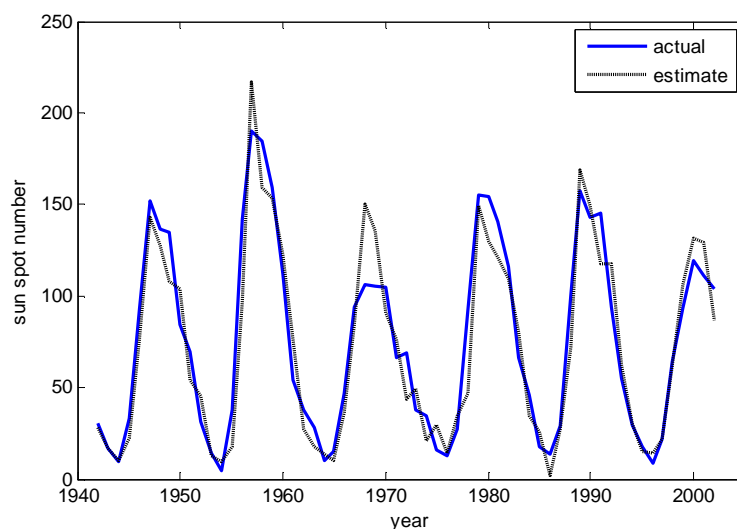
الگوریتم	FCE	LoLiMoT
میانگین مربع خطای نرمالیزه برای یادگیری	0.0650	0.1136
میانگین مربع خطای نرمالیزه برای تست	0.1090	0.1159
تعداد مدل های خطی-محلی (LLM)	5	3

نمودار تغییرات خطای یادگیری و تست به ازای تعداد نوروں های مختلف با الگوریتم خوشه یابی FCV ترکیبی در شکل 9 نمایش داده شده است. به علاوه، برای بهترین مدل به دست آمده، نمودار خروجی واقعی و خروجی تخمین زده شده و همینطور نمودار خطای تخمین بر روی داده های تست در شکل های 10 و 11 نشان داده شده اند.

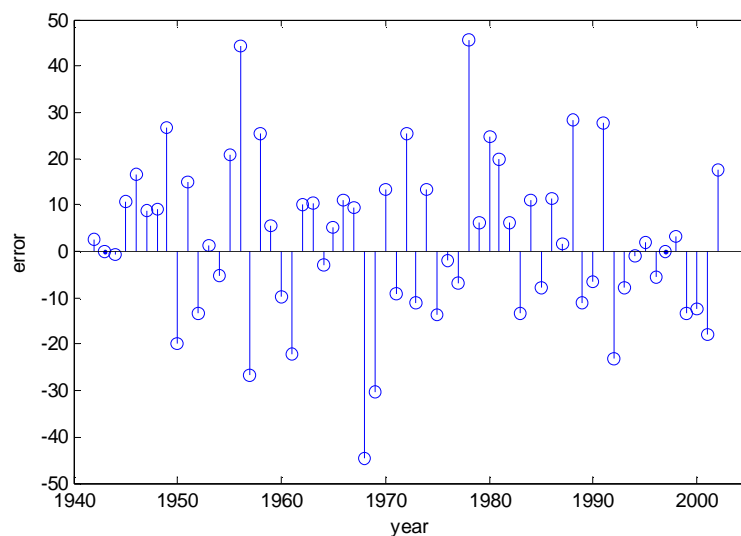


شکل 9- خطای تخمین در الگوریتم پیشنهادی به ازای تعداد نوروں (مدل های خطی-محلی) برای داده های یادگیری و تست





شکل 10- نمودار خروجی واقعی و خروجی تخمین زده شده بر روی داده ی تست



شکل 11- نمودار خطای تخمین بر روی داده تست

5.2 پیش بینی میزان نیاز آبی گیاه در شبکه آبیاری سد حسنلو

همانطور که قبلاً توضیح داده شد، برای پیش بینی میزان نیاز آبی گیاه از کمیتی به نام تبخیر-تعرق استفاده می شود. ما در این قسمت به پیش بینی سری زمانی تبخیر-تعرق پتانسیل (ET₀) (که در بخش 2 معرفی شد) به صورت دهه ای می پردازیم. بدین معنی که متوسط این کمیت در طول یک دهه به عنوان متغیر سری زمانی در نظر گرفته میشود. مقادیر این متغیر طی سالهای 1985 تا 2005 برای شبکه آبیاری سد حسنلو در جدول 4 نشان داده شده است.

جدول 4- مقادیر متوسط ET0 در مقیاس ده روزه طی 21 سال آماری

سال/دهه	1985	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005
1	0.97	2.38	2.25	1.02	1.29	1.18	1.53	1.36	1.13	1.54	1.96	1.56	2.57	1.30	1.43	1.98	1.71	1.74	2.20	1.38	1.24
2	1.61	2.03	2.50	0.83	1.38	1.163	1.14	1.46	1.01	1.54	2.19	1.11	1.80	1.35	1.44	1.22	2.06	1.91	1.68	1.93	1.66
3	1.19	1.34	2.46	1.53	1.14	1.255	1.82	1.32	2.12	1.644	2.59	1.80	1.59	1.47	2.65	1.74	1.26	1.36	1.52	2.35	1.20
4	1.59	1.947	2.26	1.47	1.20	1.453	1.68	1.43	1.33	1.847	2.06	1.81	1.81	1.81	2.43	2.33	2.14	2.71	2.92	2.66	1.23
5	2.40	2.245	2.91	0.99	1.14	1.752	2.45	2.16	2.02	2.143	2.11	2.27	2.64	1.79	3.14	2.41	2.59	2.47	1.78	2.41	1.86
6	1.89	2.627	2.59	2.88	2.22	2.142	2.57	1.50	2.207	2.523	3.21	2.69	2.22	2.31	3.43	2.99	2.55	3.30	2.14	2.85	2.67
7	1.92	3.08	3.17	2.64	2.40	2.612	2.57	2.11	2.49	2.974	3.87	3.03	2.13	3.43	4.39	2.98	3.70	4.39	2.81	3.40	2.70
8	2.88	3.592	2.49	4.10	3.07	3.148	2.40	2.92	2.34	3.484	3.24	2.24	2.02	2.83	3.62	4.26	4.43	3.43	3.23	3.41	4.16
9	3.45	4.147	2.99	3.84	3.84	3.96	3.42	3.40	4.95	4.14	4.66	2.92	3.45	3.66	3.17	3.74	4.81	4.08	3.17	4.99	4.03
10	4.30	4.727	3.98	4.11	5.07	3.72	4.35	4.06	4.29	4.69	4.41	3.48	3.26	4.78	4.29	4.31	4.45	3.74	4.29	4.11	3.85
11	5.72	5.315	4.74	4.42	6.21	5.013	4.62	4.33	5.052	5.87	5.281	3.96	4.78	4.85	5.37	5.04	5.90	3.87	4.13	4.56	5.28
12	4.03	5.893	6.58	3.98	5.83	5.719	6.08	5.39	5.646	4.67	5.829	5.43	5.71	4.58	6.31	5.65	6.11	4.91	4.25	5.14	6.16
13	6.19	6.442	6.64	6.46	6.26	6.398	6.30	2.72	6.216	5.49	6.36	5.74	5.50	5.70	6.28	5.94	5.45	5.46	6.51	5.87	5.67
14	6.45	6.946	5.94	7.19	7.07	7.03	6.78	5.18	6.744	6.46	6.857	5.54	5.98	6.41	6.89	6.67	7.51	6.43	6.98	5.22	5.89
15	7.26	7.388	7.01	7.43	7.94	7.71	7.06	5.58	7.214	7.47	7.308	8.09	7.23	7.15	8.07	8.02	8.03	8.28	5.36	6.34	7.66
16	7.67	7.756	7.37	7.95	7.82	8.55	8.06	6.85	7.612	7.81	7.697	7.72	7.31	8.26	8.16	8.58	8.77	7.99	7.43	8.80	8.01
17	8.34	8.037	7.05	7.74	8.58	8.33	7.70	6.98	7.925	8.66	7.70	8.29	8.47	8.56	8.54	8.29	8.42	8.77	8.10	8.34	9.02
18	8.53	8.222	8.54	7.32	8.57	8.59	7.90	8.27	8.144	6.90	8.44	8.16	8.58	9.08	8.88	8.87	9.55	8.89	8.61	8.98	8.26
19	7.70	8.306	8.21	7.64	8.77	8.17	8.69	9.26	8.262	8.55	8.46	7.94	7.93	8.55	8.72	8.68	8.77	8.71	8.36	8.46	8.67
20	8.20	8.285	8.83	8.31	8.25	7.93	8.16	8.40	8.276	8.60	8.59	8.45	8.90	8.50	7.73	7.57	7.59	8.80	9.01	8.42	9.09
21	7.61	8.16	7.80	8.08	8.12	8.61	8.05	8.75	8.186	8.27	8.79	8.05	8.99	7.43	7.96	8.36	8.18	8.40	8.20	8.58	8.29
22	7.79	7.81	7.60	7.11	8.14	8.29	7.85	7.43	7.993	8.30	8.57	8.41	8.03	7.69	7.46	8.21	8.68	9.07	8.48	8.33	8.68
23	7.67	7.83	8.20	7.36	7.57	7.91	7.30	7.48	7.705	7.76	8.35	7.94	8.26	8.34	7.10	7.85	8.25	8.65	8.35	8.78	8.15
24	6.87	7.07	7.16	7.33	7.50	7.25	7.37	7.06	7.331	7.07	8.25	7.40	8.14	7.47	8.13	8.43	8.02	8.04	7.56	7.83	7.74
25	7.02	6.85	6.06	7.11	7.39	7.09	7.22	6.86	6.881	5.66	7.17	6.84	6.74	7.25	7.62	7.32	7.48	8.26	7.57	8.52	7.36
26	6.48	6.36	6.01	6.48	6.32	6.85	6.60	6.46	6.37	6.37	6.57	6.95	6.35	6.24	6.75	6.99	6.96	7.43	6.96	6.49	6.99
27	5.43	6.28	6.01	5.97	5.93	5.994	6.00	5.78	5.815	5.41	5.77	6.16	5.99	5.98	5.75	6.00	6.41	6.79	6.04	6.55	6.58
28	5.00	4.66	3.86	5.25	5.76	5.374	4.51	5.20	5.231	4.84	5.53	5.66	5.58	5.01	5.84	5.06	5.63	5.30	5.88	6.03	5.43
29	4.33	4.01	3.57	3.53	4.81	4.743	3.43	4.77	4.638	3.89	4.34	4.23	4.03	4.93	5.35	4.24	5.23	4.39	4.83	5.00	4.99
30	3.74	3.88	2.01	4.30	3.05	3.64	4.54	4.69	4.053	4.03	4.43	4.04	3.53	4.37	3.00	3.62	4.00	4.40	4.40	3.98	4.40
31	3.47	1.93	3.30	3.57	3.13	3.84	2.72	3.27	3.494	2.77	3.35	3.06	2.49	4.31	3.17	3.10	3.70	3.33	3.22	1.91	3.34
32	2.85	1.91	3.60	2.88	2.66	3.42	3.30	3.21	2.979	2.90	3.56	2.79	2.78	3.49	3.02	2.53	2.53	2.94	2.70	2.70	2.37
33	2.47	1.78	2.99	2.50	2.38	2.49	2.89	2.56	2.524	2.38	2.90	3.21	2.44	3.07	2.07	2.06	2.04	2.15	1.76	2.45	2.85
34	2.07	1.69	1.80	2.63	1.96	2.45	1.67	2.123	2.142	1.68	1.91	1.91	2.45	2.75	2.15	1.74	1.20	1.32	1.51	1.88	3.23
35	1.39	1.62	1.68	2.14	1.534	2.28	1.50	1.84	1.845	1.77	1.70	2.34	1.83	2.91	1.70	1.89	2.03	0.96	1.89	1.53	2.62
36	1.65	1.86	1.29	1.81	1.305	0.86	1.48	1.47	1.643	1.28	2.66	2.34	1.84	2.15	2.41	1.83	2.22	1.32	1.19	1.40	1.96

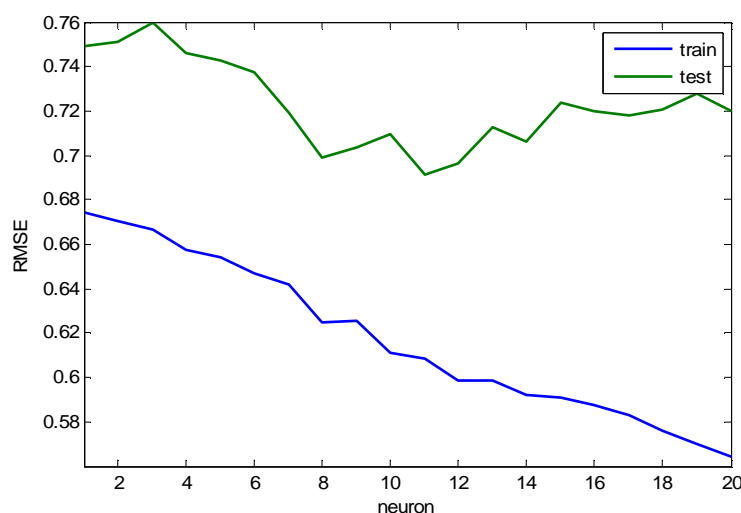


ما ابتدا روش شناسایی پیشنهادی را برای پیش بینی سال های 2000 تا 2005 با استفاده از داده های 1985 تا 1999 به کار می بریم و نتایج را با الگوریتم LoLiMoT مقایسه می کنیم. سپس سال 1992 که یک سال کم آبی بوده است را با

داده های سال های گذشته ی آن پیش بینی نموده و نتایج را با نتایج به دست آمده در [18] با استفاده از روش های دیگر مقایسه می کنیم.

5.2.1 پیش بینی سال های 2000 تا 2005

ما در این قسمت به پیش بینی کمیت تبخیر-تعرق پتانسیل طی سال های 2000 تا 2005 می پردازیم. بدین منظور، نمونه های مربوط به سال های 1985 تا 1999 به عنوان داده یادگیری و بقیه به عنوان تست در نظر گرفته می شوند. همچنین به عنوان ورودی از lag 3 اول استفاده می کنیم اگرچه که این بهترین انتخاب نیست (مثلا lag دوازدهم بسیار مهم است)، ولی بر اساس آزمایش های انجام شده در [18] به طور کلی نتایج خوبی به دست میدهد. روش پیشنهادی با الگوریتم خوشه یابی FCV ترکیبی و $\alpha = 0.9$ به کار گرفته شده و برای هر تعداد LLM خوشه یابی برای 30 بار تکرار می شود. نمودار تغییرات خطای یادگیری و تست به ازای تعداد نورون ها (LLM ها) در شکل 12 نشان داده شده است.

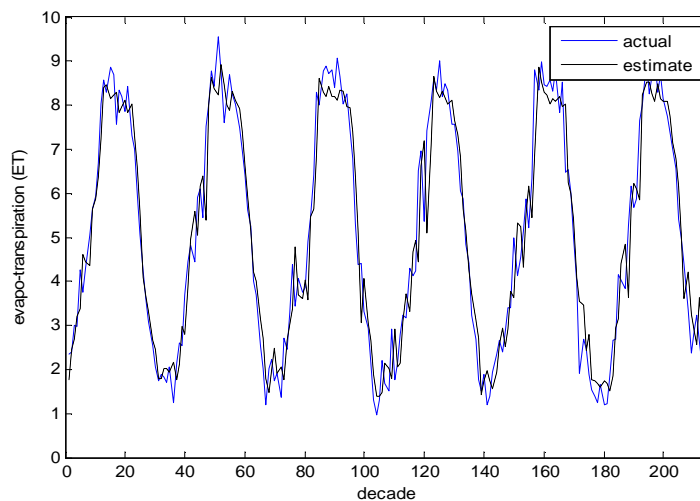


شکل 12- خطای تخمین در الگوریتم پیشنهادی به ازای تعداد نورون (مدل های خطی-محلی) برای داده های یادگیری و تست

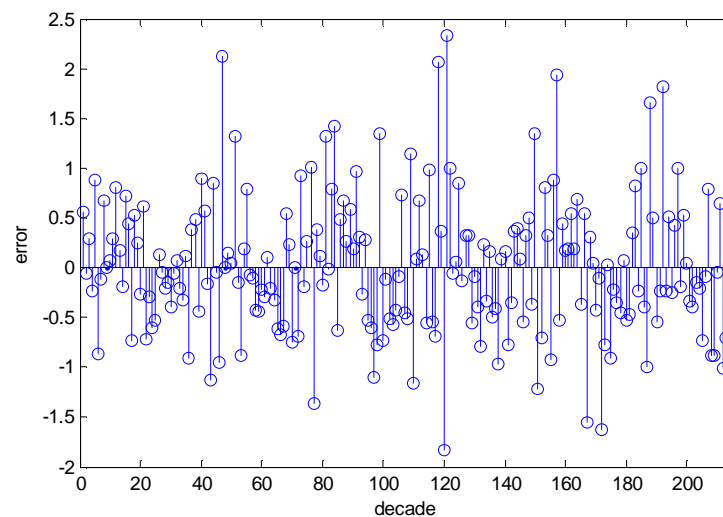
با توجه به شکل 12 بهترین نتیجه به ازای 11 مدل خطی-محلی حاصل میشود که نتایج برای این حالت در جدول 5 خلاصه شده و با الگوریتم LoLiMoT مقایسه گردیده است. علاوه بر این نمودار خروجی واقعی و خروجی تخمین زده شده و همینطور نمودار خطای تخمین بر روی داده های تست در شکلهای 13 و 14 نشان داده شده اند.

جدول 5- نتایج پیش بینی تبخیر-تعرق برای سال های 2000 تا 2005 با روش های مختلف

الگوریتم	FCE	LoLiMoT
مجذور میانگین مربع خطای یادگیری	0.5642	0.6532
مجذور میانگین مربع خطای تست	0.6912	0.7238
تعداد مدل های خطی-محلی (LLM)	11	5



شکل 13- نمودار خروجی واقعی و خروجی تخمین زده شده بر روی داده ی تست

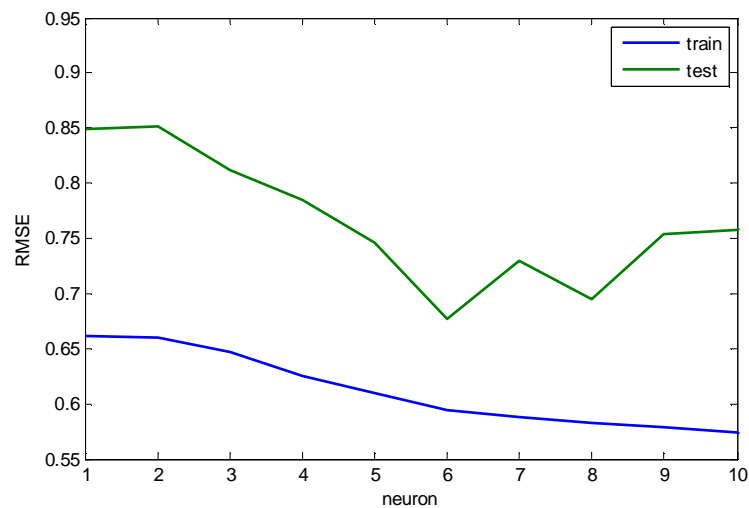


شکل 14- نمودار خطای تخمین بر روی داده تست

5.2.2 پیش بینی سال 1992

سال های 2000 تا 2005 که در قسمت قبل مورد بررسی قرار گرفتند سال های نرمالی بودند و در نتیجه پیش بینی آن ها ساده است، ولی سال 1992 نمونه ای از یک سال کم آب می باشد که پیش بینی آن از اهمیت به سزایی برخوردار است. در اینجا نیز همانند قسمت قبل 3 lag اول به عنوان ورودی انتخاب می شوند، اما داده یادگیری از سالهای 1985 تا 1991 انتخاب میشود. برای این مسئله از الگوریتم خوشه یابی FCV در شناسایی ساختار مدلهای نورو فازی استفاده کرده و خوشه یابی برای هر مدل خطی-محلی 20 بار تکرار می شود. . نمودار تغییرات خطای یادگیری و تست به ازای تعداد نورو ن ها (LLM) در شکل 15 نشان داده شده است.



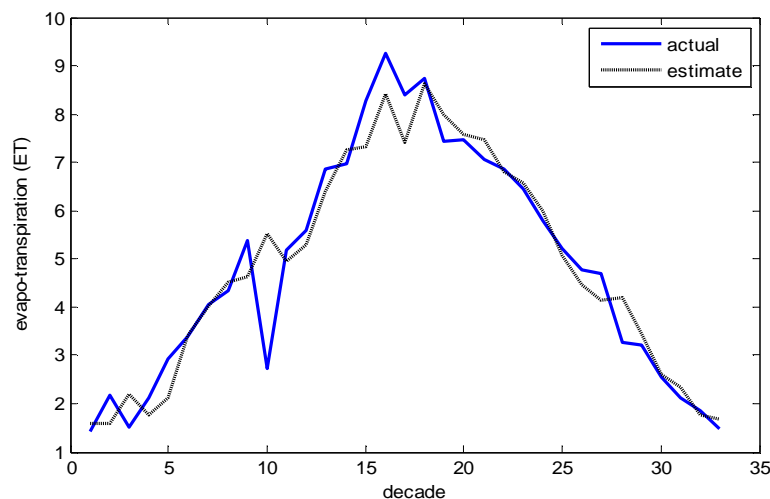


شکل 15- خطای تخمین در الگوریتم پیشنهادی به ازای تعداد نورون (مدل های خطی-محلی) برای داده های یادگیری و تست

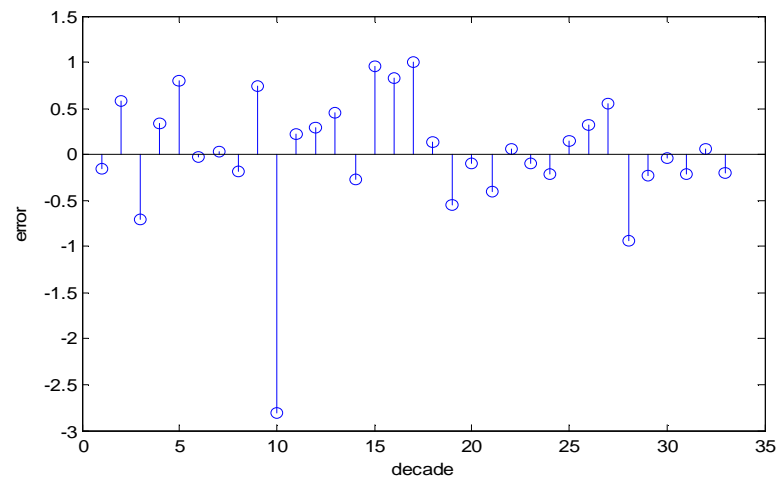
با توجه به این شکل، بهترین نتیجه به ازای 6 مدل خطی-محلی حاصل میشود که نتایج برای این حالت در جدول 6 خلاصه شده و با روش MLP و متوسط 21 سال (متوسط گیری روش سنتی پیش بینی میزان نیاز آبی گیاه می باشد) که در [18] مطرح شده اند مورد مقایسه قرار گرفته است. با توجه به این جدول، روش پیشنهادی نتایج بهتری را به دست می دهد. علاوه بر این نمودار خروجی واقعی و خروجی تخمین زده شده و همینطور نمودار خطای تخمین بر روی داده های تست در شکل های 16 و 17 نشان داده شده اند.

جدول 6- نتایج پیش بینی تبخیر-تعرق برای سال های 2000 تا 2005 با روش های مختلف

الگوریتم	FCV	MLP	متوسط 21 سال
مجذور میانگین مربع خطای تست	0.6769	0.7664	0.8050



شکل 16- نمودار خروجی واقعی و خروجی تخمین زده شده بر روی داده ی تست



شکل 17- نمودار خطای تخمین بر روی داده تست

6 نتیجه گیری

ما در این پژوهش از ایده خوشه یابی فازی با ساختار های خطی برای شناسایی ساختار مدل های نوروفازی خطی - محلی استفاده کردیم. توانایی الگوریتم ارائه شده بر روی تعدادی از مسائل معروف تخمین و شناسایی، شامل مدلسازی دو تابع غیرخطی و پیش بینی سری زمانی لکه های خورشیدی مورد آزمون قرار گرفته و با الگوریتم ها و مدل های دیگری به ویژه الگوریتم LoLiMoT مقایسه شد. به علاوه، کاربرد این الگوریتم برای پیش بینی سری زمانی تبخیر-تعرق که به عنوان معیاری در اندازه گیری میزان نیاز آبی گیاه استفاده مس شود مورد بررسی قرار گرفت. در تمامی آزمون های انجام شده، الگوریتم پیشنهادی به خطای تخمین کمتری در مقایسه با LoLiMoT دست یافت، ولی لازم به ذکر است که به دلیل محلی بودن الگوریتم خوشه یابی مورد استفاده، این عمل باید چندین بار تکرار شود و در نتیجه الگوریتم LoLiMoT بسیار سریعتر است. البته ممکن است برای مسائلی که تعداد ابعاد ورودی زیاد است و LoLiMoT به تقسیم بندی های زیاد احتیاج دارد، الگوریتم ارائه شده مفید تر باشد.

نکته ای که بسیار حائز اهمیت است و می تواند در کار های آتی مورد مطالعه قرار گیرد، استفاده از این ایده با رویکرد برخط به خوشه بندی داده ها و تبدیل روابط حاضر به روابط incremental به منظور شناسایی و رگرسیون برخط توابع و سیستم هاست. به نظر میرسد که به دلیل اینکه در حالت incremental تمامی الگوریتم ها محکوم به محلی بودن هستند، سرعت الگوریتم حاصل با الگوریتم های دیگر قابل مقایسه بوده و در عین حال ممکن است به خطای کمتری منجر شود. نکته ی دیگر اینست که ما در اینجا برای سادگی کار، بهینه سازی تابع هزینه برای FCV ترکیبی را هم برای قسمت با ساختار خطی (ترم دوم از تابع هزینه) و هم قسمت کروی در فضای تعمیم یافته ی ورودی و خروجی بررسی کردیم. در حالیکه احتمالاً بهتر است که قسمت کروی که معطوف به نزدیکی اقلیدسی داده هاست پ، از ابتدا در فضای ورودی تعریف شود.



- [1] M. Sugeno and G.T. Kang, "Structure Identification of Fuzzy Model", *Fuzzy Sets and Systems*, 28(1) 15-33, 1988.
- [2] K. M. Alexiev and O. I. Georgieva, "Improved Fuzzy Clustering for Identification of Takagi-Sugeno Model", *Proc. IEEE Int. Conf. Intelligent Systems*, Sofia, Bulgaria, 2004.
- [3] J. Abonyi, R. Babuska, and F. Szeifert, "Modified Gath-Geva Fuzzy Clustering for Identification of Takagi-Sugeno Fuzzy Models", *IEEE Trans. Systems Man Cybernetics*, 32(5) 612- 621, 2002.
- [4] T. A. Johansen and R. Babuska, "Multiobjective Identification of Takagi-Sugeno Fuzzy Models", *IEEE Trans. Fuzzy Systems*, 11(6) 847- 860, 2003.
- [5] D. Saez and R. Zuniga, "Cluster Optimization for Takagi & Sugeno Fuzzy Models and Its Application to a Combined Cycle Power Plant Boiler", *Proc. IEEE American Control Conference*, Boston, MA, USA, 2004.
- [6] O. Nelles, *Nonlinear System Identification*, Springer, 2001.
- [7] A. Kalhor, B. N. Araabi, and C. Lucas, "A new split and merge algorithm for structure identification in Takagi-Sugeno fuzzy model," in *Proc. 7th International Conference on Intelligent Systems Design and Applications*, pp. 258-261, 2007.
- [8] Trajkovic, S., Todorovic, B., Standkovic, M., 2003. Forecasting of reference evapotranspiration by artificial neural network. *ASCE Journal of Irrigation and Drainage Engineering* .129(6), 454-457.
- [9] Zanetti S.S., Sousa, E.F., Olivera, V.P.S., Almeida, F.T. and Bernardo, S., 2007. Estimating evapotranspiration using artificial neural network and minimum climatological data . *ASCE Journal of Irrigation and Drainage Engineering*. 133(2), 83-89.
- [10] Sudheer KP, Gosain AK, Ramasastri KS. 2003a. Estimating actual evapotranspiration from limited climatic data using neural computing technique. *ASCE Journal of Irrigation and Drainage Engineering*. 129(3), 214-218.
- [11] Keskin ME, Terzi O. 2006. Artificial neural network models of daily pan evaporation. *Journal of Hydrologic Engineering*, ASCE 11(1): 65-70.
- [12] J. C. Bezdek, C. Coray, R. Gunderson, J. Watson, "Detection and Characterization of Cluster Substructure I. Linear Structure: Fuzzy c-Lines," *SIAM Journal on Applied Mathematics*, Vol. 40, No. 2 (Apr., 1981), pp. 339-357.
- [13] J. C. Bezdek, C. Coray, R. Gunderson, J. Watson, "Detection and Characterization of Cluster Substructure II. Fuzzy c Varieties and Convex Combinations Thereof," *SIAM Journal on Applied Mathematics*, Vol. 40, No. 2 (Apr., 1981), pp. 358-372.
- [14] J-S. R. Jang, "ANFIS: Adaptive-Network-Based Fuzzy Inference System", *IEEE Trans. Systems Man Cybernetics*, 23(3) 665-685, 1993.
- [15] M. Sugeno and G.T. Kang, "Structure Identification of Fuzzy Model", *Fuzzy Sets and Systems*, 28(1) 15-33, 1988.
- [16] H. Takagi and I. Hayashi, "NN-driven fuzzy reasoning", *International Journal of Approximate Reasoning*, 5(3) 191-212, May 1991.
- [17] M. M. Rezaei Yousefi, M. Mirmomeni, and C. Lucas, "Input variable selection using mutual information for neuro fuzzy modeling with application to time series forecasting," in *Proc. Int. Joint Conf. on Neural Networks*, Orlando, Florida, USA, 2007.

[18] امین بمانی، برآورد تبخیر-تعرق با استفاده از داده های اقلیمی همزمان در راستای بهبود مدیریت آب در شبکه های آبیاری، پایان نامه ی کارشناسی ارشد، دانشکده کشاورزی دانشگاه تهران، 1387.

