



Prediction of Evaporation Using Chaos Theory and Artificial Intelligence in Dry Lands (Case Study: Semnan Province)

F. Hoshmandzadeh¹, M.R. Yazdani^{2*}, and S.F. Mousavi³

Abstract

Evaporation is one of the important phenomena of the hydrological cycle and its prediction is essential for water management, planning and conservation. Since chaos theory deals with the study of dynamic systems, in this research the prediction of the evaporation process was carried out using the combination of chaos theory and intelligent models, including support vector machine, decision tree, group learning, and Gaussian process. Data of the Semnan synoptic station during the period of 1995-2019 was selected. The optimal values of delay and mutual information were respectively obtained as 18 and 9 using false nearest neighbor methods in order to reconstruct the variable phase space of evaporation. According to different combinations of variables, the most optimal response of all models was determined for the combination of all parameters, and the two factors of evaporation and temperature had the greatest impact on the prediction. In general, the support vector machine model with $R^2 = 85.5$ and $MAE = 1.4$ had the best performance followed by the methods of Gaussian process, group learning and decision tree method as next bests. The combined use of chaos theory along with intelligent algorithms has a good ability to estimate evaporation.

Keywords: Evaporation Forecasting, Intelligent Models, Chaos Theory, Water Conservation.

Received: November 19, 2023

Accepted: April 23, 2024

پیش‌بینی تبخیر با استفاده از تئوری آشوب و هوش مصنوعی در مناطق خشک (مطالعه موردی: استان سمنان)

فاطمه هوشمندزاده^۱، محمدرضا یزدانی^{۲*} و سید فرهاد موسوی^۳

چکیده

تبخیر از پدیده‌های مهم چرخه آب‌شناختی است و پیش‌بینی آن در مدیریت، برنامه‌ریزی و حفظ آب ضروری است. از آنجائی که تئوری آشوب به مطالعه سیستم‌های دینامیکی می‌پردازد، لذا در تحقیق حاضر، پیش‌بینی فرآیند تبخیر با استفاده از ترکیب تئوری آشوب و مدل‌های هوشمند شامل ماشین بردار پشتیبان، درخت تصمیم، یادگیری گروهی، و فرآیند گوسی انجام شد و داده‌های ایستگاه سینوپتیک سمنان طی دوره زمانی ۱۹۹۵-۲۰۱۹ انتخاب شد. مقادیر بهینه میزان تأخیر و اطلاعات متقابل با استفاده از روش‌های نزدیکترین همسایه نادرست به منظور بازسازی فضای فاز متغیر تبخیر، به ترتیب برابر با ۱۸ و ۹ بدست آمد. با توجه به ترکیب‌های متفاوتی از متغیرها، بهینه‌ترین پاسخ همه‌ی مدل‌ها برای ترکیب کلیه پارامترها مشخص شد و دو عامل تبخیر و دما بیشترین تأثیر را بر پیش‌بینی داشتند. به طور کلی، مدل ماشین بردار پشتیبان با $R^2 = 85.5$ و $MAE = 1.4$ بهترین کارایی را داشت و سپس، به ترتیب روش‌های فرآیند گوسی، یادگیری گروهی و روش درخت تصمیم کارایی مناسبی داشتند. استفاده ترکیبی تئوری آشوب به همراه الگوریتم‌های هوشمند از قابلیت خوبی برای تخمین تبخیر برخوردار است.

کلمات کلیدی: پیش‌بینی تبخیر، مدل‌های هوشمند، نظریه آشوب، حفاظت آب.

تاریخ دریافت مقاله: ۱۴۰۲/۸/۲۸

تاریخ پذیرش مقاله: ۱۴۰۳/۲/۴

1- Ph.D. Graduate in Desert Management and Control, Faculty of Desert Studies, Semnan University, Semnan, Iran.

2- Associate Professor, Faculty of Desert Studies, Semnan University, Semnan, Iran. Email: m_yazdani@semnan.ac.ir

3- Professor, Faculty of Civil Engineering, Semnan University, Semnan, Iran.

*- Corresponding Author

Doi: [10.22034/IWRR.2024.426028.2721](https://doi.org/10.22034/IWRR.2024.426028.2721)

۱- دانش‌آموخته دکتری مدیریت و کنترل بیابان، دانشکده کویرشناسی، دانشگاه سمنان، سمنان، ایران.

۲- دانشیار گروه بیابانزادایی، دانشکده کویرشناسی، دانشگاه سمنان، سمنان، ایران.

۳- استاد دانشکده مهندسی عمران، دانشگاه سمنان، سمنان، ایران.

*- نویسنده مسئول

بحث و مناظره (Discussion) در مورد این مقاله تا پایان تابستان ۱۴۰۳ امکان‌پذیر است.



مرسین و آنتالیا با ورودی محلی و داده‌های خروجی داشت. در تحقیقی برای تخمین میزان تبخیر از سطح دریاچه ارومیه از تئوری آشوب و شبکه عصبی در یک دوره ۴۰ ساله استفاده شد. بررسی شاخص‌های آشوب مانند بدلیاپانف مبین رفتار آشوبناک سری زمانی داده‌های تبخیر بوده است. از طرف نتایج صحت‌سنجی مدل حاکی از این است که مدل شبکه عصبی مصنوعی خطای کمتری نسبت به تئوری آشوب داشته است (Farzin et al., 2017). از جمله پژوهش‌های مشابه دیگر در این زمینه می‌توان به تحقیقات (Kumar Goyal et al., 2017; Malik et al., 2017; Ferreira et al., 2019, Ayodel et al., 2019) اشاره کرد. Khular and Singh (2022) یک مدل مبتنی بر یادگیری عمیق (DP) برای پیش‌بینی عوامل کیفیت آب رودخانه یامونا، هند معرفی کردند. داده‌ها شامل نمونه‌های ماهانه عوامل مختلف کیفیت آب به مدت ۶ سال (۲۰۱۹-۲۰۱۳) در چندین مکان در منطقه دهلی جمع‌آوری شد. نتایج تجربی نشان داد که مقادیر پیش‌بینی‌شده مدل و مقادیر واقعی مطابقت نزدیکی دارند و می‌توانند روند آینده را نشان دهند. عملکرد مدل با تکنیک‌های مختلف از جمله SVR، جنگل تصادفی (RF) و شبکه عصبی مصنوعی مقایسه شد. تجزیه و تحلیل تجربی با اندازه‌گیری سطوح اکسیژن مورد نیاز بیولوژیک و شیمیایی انجام و نتایج نشان داد که مدل پیشنهادی دارای دقت خوبی در پارامترهای کیفی آب دارد.

(La Fuente et al., 2022) پیش‌بینی تبخیر برای دریاچه کینرت را با استفاده از چند مدل انجام دادند. در این مطالعه، یک رویکرد چند مدلی به همراه یک الگوریتم آبرودینامیکی مورد بررسی قرار گرفت و نتایج نشان داد که رویکرد چند مدلی با $RMSE=0.19$ میلیمتر و $R^2 = 0.92$ عملکرد بهتری داشت. El Bilali et al. (2023) در ایستگاه هواشناسی سما در مراکش تبخیر را با استفاده از ۵ مدل، از جمله یادگیری ماشینی (ML) مورد پیش‌بینی قرار دادند. یافته‌ها نشان داد مدل‌های ML برای کاربرد در فرایندهای مرتبط با تبخیر واقعی در مناطق خشک مناسب هستند.

Rocha et al. (2023) مطالعه‌ای با هدف ارزیابی تغییرپذیری مکانی-زمانی نرخ تبخیر در سطح آب مخزن پنتیکوست، واقع در منطقه نیمه خشک برزیل، با استفاده از مدل هیدرودینامیکی Delft3D و تکنیک سنجش از دور انجام دادند. با توجه به نتایج توزیع مکانی، برای دوره مرطوب، الگوهای تبخیر شبیه به الگوهای سنجش از دور بود، در حالی که در دوره خشک، سنجش از دور در مقایسه با Delft3D الگوی تبخیر پایدارتری ارائه کرد. رویکرد نوآورانه ارائه شده در مطالعه مذکور می‌تواند برای درک بهتر دینامیک تبخیر در آب‌های سطحی و بهینه‌سازی مکان سازه‌های تبخیر استفاده شود.

تبخیر، از عمده‌ترین اجزای چرخه هیدرولوژی است که تخمین درست آن در طراحی و مدیریت سیستم‌های آبیاری، مطالعات منابع آب و موارد مشابه دیگر از اهمیت زیادی برخوردار است (Alsumaiei, 2020). میزان آبی که از مخازن سدها در مقیاس جهانی به صورت تبخیر هدر می‌رود از مجموع آب مورد نیاز صنعت و شرب بیشتر است. با این حال، این حجم زیاد آب تلف شده به دلیل پیچیدگی‌هایی که در برآورد این تلفات وجود دارد، فقط در شیوه‌های مدرن مدیریت منابع آب مورد توجه قرار می‌گیرد. از دست رفتن آب مخازن در مناطق گرم و خشک می‌تواند به اندازه‌ای قابل توجه باشد که اثرات مثبت نگهداشت آب در پشت مخازن را خنثی کند (Zhang et al., 2017). اخیراً با توجه به پیشرفت‌های فناوری و مدل‌های یادگیری ماشین بدلیل توانایی آنها در مدل‌سازی روابط غیر خطی پیچیده، در تخمین تبخیر از آنها استفاده شده است (Sik Kim et al., 2023; Zhao and Gao, 2019). در پژوهشی کاربرد شبکه‌های عصبی مصنوعی و مدل‌های اقلیمی (پنمن پرستیلی - تیلور و استفان استوارت) در تخمین تبخیر از مخزن آبی در لیبی مورد بررسی قرار گرفت. بر طبق یافته‌های حاصله در مقایسه روش‌های اقلیمی، روش استفان استوارت با استفاده از داده‌های تابش خورشیدی بهترین عملکرد را از خود نشان داد (Benzaghta, 2014). میزان تبخیر ماهانه دریاچه ارومیه طی یک دوره ۴۰ ساله با استفاده از تئوری آشوب مورد بررسی قرار گرفت. بعد همبستگی و تابع لیاپانوف برای تشخیص وجود رفتار آشوبناک و تصادفی استفاده شد. بعد همبستگی برابر $2/47$ بدست آمد و از طرفی مقدار مثبت تابع لیاپانوف به همراه طیف توان فوریه مبین رفتار آشوبناک تبخیر آب بوده است. در این تحقیق از الگوریتم نزدیکترین همسایه کاذب برای تخمین تبخیر با استفاده از تئوری آشوب استفاده شده است. صحت‌سنجی داده‌ها برای یک دوره ۱۰ ساله مبین کارایی بالای این روش بوده است (Farzin et al., 2012). Ozlem and Kasim (2014) به پیش‌بینی تبخیر بر اساس مدل ماشین بردار پشتیبان (SVM) همراه با رویکرد آشوب^۲ پرداختند. برای تهیه داده‌های ورودی SVM، بازسازی فضای حالت با استفاده از روش‌های تعبیه شده سری زمانی یک و چند متغیره انجام شد. نتایج نشان داد که پیش‌بینی‌های مبتنی بر SVM به ترتیب ۸۳ و ۹۷ درصد موفق به کسب ضریب تعیین بیشتر شده‌اند. Kisi (2015) مدل‌سازی تبخیر را با استفاده از داده‌های دو ایستگاه مرسین و آنتالیا در منطقه مدیترانه‌ای ترکیه انجام داد و دقت ماشین بردار پشتیبان حداقل مربعات، اسپلاین رگرسیون چند متغیره (MVRS)^۳ و مدل درخت تصمیم (M5Tree) در این تحقیق بررسی شد. مدل ماشین بردار پشتیبان عملکرد بهتری نسبت به مدل‌های MARS و M5Tree در برآورد پارامترهای ایستگاه

در استان سمنان تغییرات ماهانه دما زیاد است. آب و هوای شهر سمنان در تابستان گرم و در زمستان نسبتاً سرد است. متوسط حجم بارش سالانه در استان سمنان ۱۱/۵ میلیارد متر مکعب برآورد شده که ۹/۲ میلیارد متر مکعب آن (حدود ۸۰ درصد) مستقیماً از طریق تبخیر از دسترس خارج می‌شود و در نتیجه، به منابع آب زیرزمینی نفوذ نکرده و در تعادل بخشی نیز مؤثر نیست، که جزو مشکلات منابع آب استان سمنان است. بنابراین با عنایت به عدم وجود اطلاعات جوی از آینده، برای برنامه‌ریزی منابع آب و مدیریت آب نیاز است تا به گونه‌ای تبخیر در دوره‌های زمانی متفاوت پیش‌بینی شود. از طرفی در برخی تحقیقات انجام گرفته به ماهیت تبخیر و عوامل مؤثر بر آن کمتر توجه شده است. در این تحقیق از تئوری آشوب برای بررسی میزان آشوبناکی تبخیر استفاده شد و پس از بررسی ماهیت دینامیک آن، از مدل‌های مبتنی بر هوش مصنوعی برای تخمین تبخیر استفاده شد. همچنین، فرایند گوسی نیز برای بهبود تخمین مورد استفاده قرار گرفت. به طور کلی، نتایج حاصل از پژوهش حاضر می‌تواند به منظور پیش‌بینی‌های مناسب پدیده تبخیر و به تبع آن برنامه‌ریزی‌های دقیق و مناسب جهت کنترل پدیده مذکور مورد استفاده قرار گیرد. کاربرد این یافته می‌تواند در امر مدیریت بهینه منابع آب با توجه به اثرات محتمل تغییر اقلیم سودمند باشد.

۲- مواد و روش‌ها

۲-۱- منطقه مورد مطالعه

استان سمنان دارای میانگین سالانه بارندگی ۱۴۰ میلی‌متر و متوسط حجم بارش سالانه در استان سمنان ۱۱/۵ میلیارد متر مکعب برآورد شده است. دمای میانگین سالانه ۱۷/۷ درجه سلسیوس است. تغییرات ماهانه دما در استان سمنان زیاد است و بیشترین و کمترین دما به ترتیب مربوط به ماه‌های تیر و دی است. آب‌وهوای شهر سمنان در

تابستان گرم و در زمستان نسبتاً سرد است. در مقیاس ماهانه، بیشترین سرعت بادها مربوط به ماه‌های خرداد و مرداد، با میزان ۴/۳ متر بر ثانیه، است. در این مطالعه، اطلاعات تبخیر از تست تبخیر کلاس A ایستگاه سینوپتیک سمنان (شکل ۱)، دارای مختصات جغرافیایی به طول "۲۵'۰۰" ۵۳° و عرض "۳۶'۰۰" ۳۵° و ارتفاع ۱۱۲۷ متری از سطح دریا، در بازه زمانی اول ژانویه ۱۹۹۵ تا ۳۱ اکتبر ۲۰۱۹ دریافت و به منظور ارزیابی عملکرد ۴ روش هوش مصنوعی شامل ماشین بردار پشتیبان (SVR)، درخت تصمیم (DT)^۱، روش ترکیبی (EL)^۲ و رگرسیون فرایند گوسی (GP)^۳ به همراه تئوری آشوب برای مدل‌سازی تبخیر روزانه استفاده شد. این ایستگاه نزدیک‌ترین ایستگاه سینوپتیک به محل پژوهش می‌باشد که در فاصله ۲/۳۹ کیلومتری از آن واقع شده است. دادگان ایستگاه شامل اطلاعات میانگین دمای هوا، میانگین سرعت باد، میانگین فشار، میانگین رطوبت نسبی، ساعت آفتابی، بارندگی و تبخیر بود. در جدول ۱ مقادیر آماری داده‌ها در ایستگاه سینوپتیک سمنان و همچنین سری زمانی داده‌های تبخیر ماهانه در شکل ۲ نشان داده شده است.

۲-۲- روش تحقیق

برای توصیف رفتار دینامیکی پدیده تبخیر از بازسازی فضای فاز استفاده و برای آموزش مدل‌های هوشمند پیش‌بینی میزان تبخیر استفاده شد. با استفاده از الگوریتم نزدیک‌ترین همسایه نادرست، بعد بهینه و با کمک تابع اطلاعات متقابل، تأخیر زمانی بهینه، تخمین زده شد. با استفاده از مقادیر بعد تعبیه و زمان تأخیر، فضای فاز بازسازی شده و از این اطلاعات برای پیش‌بینی مقادیر تبخیر توسط مدل‌های ماشین بردار پشتیبان، درخت تصمیم، رگرسیون فرایند گاوسی و روش ترکیبی استفاده شد.

Table 1- Statistical data values in Semnan synoptic station from January 1, 1995 to October 31, 2019

جدول ۱- مقادیر آماری داده‌ها در ایستگاه سینوپتیک سمنان از تاریخ زمانی اول ژانویه ۱۹۹۵ تا ۳۱ اکتبر ۲۰۱۹

Statistical value	Evaporation (mm)	Temperature (Celsius)	Wind speed (m/s)	Pressure (Pascal)	Relative humidity (%)	Sunshine (hours)	Rainfall (mm)
Average	9	23.5	2.1	889	62	9.2	0.3
Standard deviation	4.84	8	1.4	4.81	15.6	3	1.61
Variance	23.53	64	2	23.2	244	9	2.6
Maximum	25	37.8	0	870	93.2	14.2	36.2
Minimum	0	1.3	12.25	902.7	6.5	0	0
Number of data	6460	6460	6460	6460	6460	6460	6460



Fig. 1- Location of Semnan synoptic station
 شکل ۱- موقعیت ایستگاه سینوپتیک سمنان

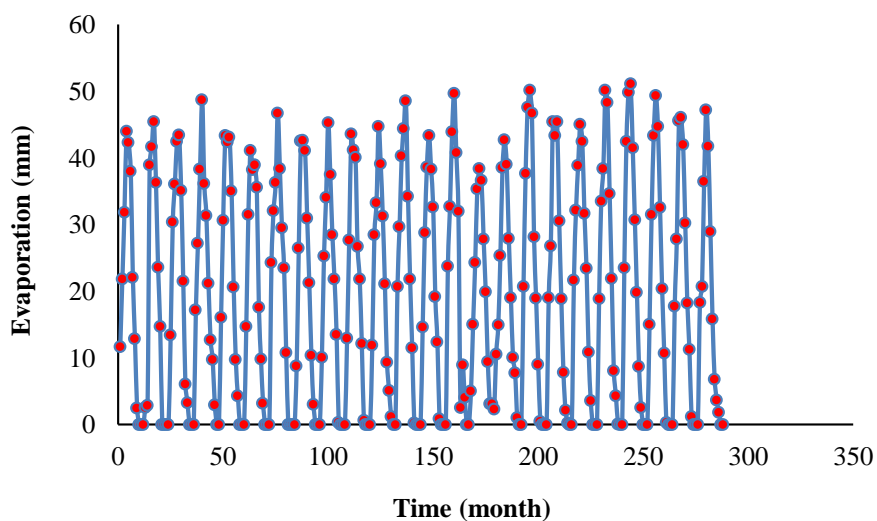


Fig. 2- Time series of monthly evaporation in Semnan station
 شکل ۲- سری زمانی تبخیر ماهانه در ایستگاه سمنان

۲-۱-۲-۱- تئوری آشوب

انتخاب ابعاد فضای فاز نیز از اهمیت خاصی برخوردار است؛ زیرا اگر بعد خیلی کوچک انتخاب شود، هندسه فضای فاز کاملاً پیچ و تاب نمی‌خورد و اگر بعد خیلی بزرگ اتخاذ شود، نیاز به محاسبات بالایی خواهد بود و مشکلات گرد کردن یا خطای ابزاری افزایش می‌یابد. نشان داده شده است که ابعاد بخش‌های مختلف سیگنال‌های غیرایستاد ممکن است با هم متفاوت باشند. از طرف دیگر، نویز تمایل دارد بعد را افزایش دهد. در نتیجه، انتخاب بعد به اندازه کافی بزرگ، که شامل تمامی دینامیک‌های مرتبط در حضور نویز باشد، اهمیت بسزایی دارد. روش نزدیکترین همسایه نادرست (FNN)^{۱۵} که بر اساس بیشترین فاصله داده‌های هر بعد کار می‌کند، می‌تواند روش مناسبی برای انتخاب بعد بهینه فضای فاز بازسازی شده باشد.

۲-۱-۲-۲- نمودار بازگشتی و اندازه‌گیری میزان قطعیت سیستم

حالات بازگشتی یک ویژگی اساسی یک سیستم دینامیکی قطعی است. نمودارهای بازگشتی می‌توانند ویژگی بازگشتی یک سیستم دینامیکی قطعی را با نشان دادن رفتار وابسته اربیت‌های x_i در فضای فاز تشریح نمایند. گام کلیدی در یک نمودار بازگشتی محاسبه ماتریس R با ابعاد $N \times N$ به صورت معادله ۳ است.

$$R_{i,j}(\varepsilon) = \Theta(\varepsilon - \|x_i - x_j\|), \quad i, j = 1, \dots, N \quad (2)$$

در این معادله ε فاصله قطع پیش‌فرض، $\| \cdot \|$ نرم اقلیدسی و $\Theta(x)$ یک تابع پله است. بردار فضای حالت x_i می‌تواند توسط روش تأخیر زمانی و بر اساس بردار مشاهدات u_i بازسازی شود ($x_i = (u_i, u_{i+\tau}, \dots, u_{i+(m-1)\tau})$). فاصله قطع ε ، یک کره به مرکز x_i تعریف می‌کند و اگر x_j داخل این کره قرار گیرد (یعنی حالت نزدیک به x_i است)، مقدار $R_{i,j} = 1$ شده و در غیر این صورت $R_{i,j} = 0$ خواهد شد. مقدار دودویی $R_{i,j}$ می‌تواند به سادگی با رنگ‌های سپید و سیاه تصویرسازی شود. بنابراین، نمودار بازگشتی به عنوان روشی برای بیان حرکت تراژکتوری در فضای فاز با ابعاد بالا می‌تواند مورد توجه قرار گیرد. مزیت دیگر نمودار بازگشتی این است که برای آنالیز داده‌های غیرایستا و کوتاه مناسب است. در نهایت، نمودار بازگشتی مشخصات الگوهای یک سیستم دینامیکی مقیاس-کوچک و مقیاس-بزرگ را بیان می‌کند.

به منظور شناسایی ویژگی‌های نمودار بازگشتی، برخی اندازه‌گیری‌های پیچیدگی که ساختارهای مقیاس-کوچک در نمودار بازگشتی را کمی می‌کند و آنالیزهای کمی بازگشتی (RQA)^{۱۵} نامیده می‌شود، پیشنهاد می‌گردد. این اندازه‌گیری‌ها بر اساس چگالی نقطه بازگشتی و ساختارهای خط قطری موجود در نمودار بازگشتی است. ساده‌ترین

نظریه آشوب به مطالعه سیستم‌هایی می‌پردازد که در نگاه اول به نظر می‌رسد رفتار تصادفی دارند. اما در واقع همین سیستم تحت حاکمیت قوانین مشخصی است و یا به عبارتی در هر بی‌نظمی، نظم نهفته است. چنین سیستمی به شرایط اولیه بسیار حساس است، به گونه‌ای که ورودی‌های ظاهراً ناچیز و دلخواه، قادرند تأثیرات شگرفی بر سیستم داشته باشند. به چنین سیستم‌هایی، سیستم‌های آشوبناک گفته می‌شود. پس ناپایداری، رفتار غیر پریودیک، سیستم‌های قطعی و غیر خطی بودن همراه با هم از مشخصات یک سیستم آشوبناک اند (Taghipour, 2023). نظریه آشوب برای اولین بار در سال ۱۹۶۵ توسط محقق به نام Edward Lawrence در هواشناسی به کار برده شد (Kocak et al., 2007). اولین گام در آنالیزهای غیرخطی یک سیگنال با کمک تئوری دینامیک غیرخطی، بازسازی فضای فاز تراژکتوری سیگنال است (Weber and Ziblot, 2005).

۲-۱-۲-۲- بازسازی فضای فاز^{۱۴}

اولین گام در آنالیزهای غیرخطی یک سیگنال با کمک تئوری دینامیک غیرخطی، بازسازی فضای فاز تراژکتوری سیگنال است. ابزار اولیه تأخیر زمانی (τ) به این منظور به کار گرفته می‌شود که یک سری زمانی اسکالر مانند u_1, u_2, \dots, u_L را به یک فضای m بعدی تبدیل کند. معادله ۲ بسط دهد.

$$x_k = (u_k, u_{k+\tau}, \dots, u_{k+(m-1)\tau}) \quad (1)$$

در این معادله τ ، $k = 1, 2, \dots, L - (m-1)\tau$ تأخیر زمانی و m ابعاد فضای بازسازی شده است ($m \geq 2$). این روش بسیار ساده به نظر می‌رسد ولی اولین مشکل انتخاب تأخیر زمانی (τ) و حداقل فضای (m) است که سری زمانی باید به آن بسط داده شود (Weber and Ziblot, 2005).

در این روش، فرض بر آن است که داده نامحدود و بدون نویز است که در این حالت تأخیر زمانی می‌تواند به صورت دلخواه انتخاب شود. ولی مشکلی که در دادگان واقعی وجود دارد این است که همیشه سری‌های زمانی محدود و نویزی هستند و در انتخاب این تأخیر زمانی باید دقت زیادی شود. اگر تأخیر خیلی کوچک انتخاب شود، جاذب بازسازی شده در راستای یک خط فشرده می‌شود و حشو اطلاعات اتفاق می‌افتد. اگر تأخیر خیلی بزرگ باشد، دینامیک جاذب ممکن است به صورت تصادفی قطع شود و بی‌ارتباطی شکل گیرد. یکی از روش‌های معمول برای انتخاب زمان تأخیر مناسب، بر اساس آشکارسازی اولین کمینه محلی تابع اطلاعات متقابل است. در واقع، این اولین کمینه تأخیری را محاسبه می‌کند که سیگنال $u_{t+\tau}$ بیشترین اطلاعات را به دانشی که نسبت به u_t موجود است، اضافه می‌کند.

اندازه‌گیری، نرخ بازگشتی (RR)^{۱۶} است (معادله ۳) که تعداد نقاط سیاه را در نمودار بازگشتی نشان می‌دهد و درصد نقاط بازگشتی موجود در محدوده ϵ را بازگو می‌کند. در معادله ۴، N تعداد نقاط بر روی تراژکتوری موجود در فضای حالت است.

$$RR(\epsilon) = \frac{1}{N^2} \sum_{i,j=1}^N R_{i,j}(\epsilon) \quad (3)$$

اندازه‌گیری قطعی (DET) بر اساس خطوط قطری است. این پارامتر سهم نقاط بازگشتی را در ساختارهای خط قطری نشان می‌دهد. بخش‌های خط قطری دارای کمترین میزان طول هستند. این اندازه‌گیری بازگوکننده میزان الگوهای قطعی و پربودیک در دینامیک سیستم است. سیگنال‌های پربودیک دارای خطوط قطری خیلی طولانی هستند و این در حالی است که سیگنال‌های آشوبگونه خطوط قطری کوتاهی دارند. سیگنال‌های تصادفی حتی اگر پارامتر شعاع خیلی بزرگ انتخاب شود، هیچ نوع خطوط قطری ندارند. محاسبه پارامتر قطعیت در معادله ۵ آورده شده است. در این معادله، $P^\epsilon(l)$ توزیع فراوانی ساختارهای قطری با طول l در نمودار بازگشتی است و N تعداد خطوط قطری در نمودار بازگشتی است. l_{min} مقدار آستانه است که از خطوط قطری تشکیل شده از حرکت تانژانتی تراژکتوری در فضای فاز مشتتی شده و در این مطالعه، مقدار ثابت ۲ در نظر گرفته شده است.

$$DET = \frac{\sum_{l=l_{min}}^N l P^\epsilon(l)}{\sum_{i,j} R_{i,j}} \quad (4)$$

پارامتر آنتروپی (ENTR)^{۱۸} (معادله ۶)، معرف آنتروپی شانونی توزیع فراوانی طول‌های خط قطری است. این پارامتر پیچیدگی ساختار قطعی را در یک سیستم دینامیکی اندازه‌گیری می‌کند. ساختار قطعی پیچیده‌تر، مقدار ENTR بزرگ‌تری را موجب خواهد شد.

$$ENTR = - \sum_{l=l_{min}}^N p(l) \ln p(l), \quad p(l) = \frac{P(l)}{\sum_{l=l_{min}}^N P(l)} \quad (5)$$

پارامتر خط پیشینه (LMAX)^{۱۹} (معادله ۷)، به سادگی اندازه طولانی‌ترین خط قطری (از قطر اصلی مستثنی است) را در نمودار بازگشتی بیان می‌کند. این شاخص به صورت معکوس با نمای لیاپانوف مثبت^{۲۰} ارتباط دارد. نمای لیاپانوف مثبت بازگوکننده واگرایی تراژکتوری‌های نزدیک به هم و شاخصی کمی برای بیان دینامیک آشوبناک است. از این رو، مقدار کوچک LMAX معرف رفتار آشوبی بیشتر در سیگنال است.

$$LMAX = \text{length of longest diagonal line} \quad (6)$$

میزان فاصله قطع (ϵ) در نمودار بازگشتی یک پارامتر ویژه‌ای است. اگر خیلی بزرگ انتخاب شود، در این صورت تقریباً هر نقطه‌ای همسایه نقاط دیگر خواهد بود و این باعث اشباع شدن نمودار بازگشتی که

شامل نقاط غیر مرتبط است خواهد شد. اگر خیلی کوچک انتخاب شود، دیگر نقاط بازگشتی موجود نخواهد بود و اطلاعات سیستم مورد نظر از بین خواهد رفت (Wallot and Mønster, 2018).

۲-۲-۲- تخمین تبخیر با مدل‌های پیش‌بینی

دادگان معرفی شده به دو بخش دادگان آموزش (کالیبراسیون) و دادگان آزمون تقسیم شد. دادگان آزمون برای ارزیابی مدل‌های پیش‌بینی‌کننده میزان تبخیر مورد استفاده قرار گرفت. به این منظور، ۸۰ درصد از دادگان به عنوان آموزش (۵۰۱۴) و ۲۰ درصد مابقی برای آزمون (۱۲۹۲)، به صورت تصادفی، انتخاب شدند. سپس، با استفاده از روش‌های ماشین بردار پشتیبان (SVR)، درخت تصمیم (DT)، روش ترکیبی (EL) و رگرسیون فرآیند گوسی (GP) مدل‌سازی صورت گرفت. نحوه تشکیل متغیرهای ورودی داده‌های آزمون و آموزش به مدل‌های تخمین‌گر، در جدول ۲ آورده شده است.

۲-۲-۲-۱- ماشین بردار پشتیبان

مدل ماشین بردار پشتیبان (SVM) اولین بار توسط Wapnik (1995) برای حل مسائل طبقه‌بندی اطلاعات و طبقه‌بندی مجموعه داده‌ها مطرح شده و سپس برای حل مسائل رگرسیون غیرخطی و تخمین چگالی توسعه داده شده است (Kumar, 2017). در روش SVM ابتدا کل داده‌ها به دو دسته آموزش و آزمون جدا می‌شوند: هر نمونه‌ای از داده‌های مجموعه آموزش دارای یک مقدار هدف x و چندین ویژگی d (ویژگی یا متغیر مشاهداتی) می‌باشد. هدف در روش SVM ایجاد مدلی بر اساس داده‌های آموزش است؛ به طوری که بتوان مقادیر هدف داده‌های آزمون (x) را از روی ویژگی‌های داده‌های آزمون (d) پیش‌بینی کرد (Wang, 2013).

$$f(x) = w\phi(x) + b \quad (7)$$

که w بردار وزن، b ثابت بایز یا خطای مدل و $\phi(x)$ یک تابع انتقال غیر خطی از ویژگی‌های ورودی x است. متغیرهای w و b را می‌توان با بهینه‌سازی تابع خطا رابطه (۹) پیش‌بینی نمود (Wang, 2013).

$$\frac{1}{2} \|w\|^2 + C \sum_{i=1}^N (\xi_i + \xi_i^*) \quad (8)$$

۲-۲-۲-۲- درخت تصمیم

روش درخت تصمیم (DT) یک روش سلسله‌مراتبی یا چند مرحله‌ای است که در آن به صورت بازگشتی مجموعه داده‌ها به روش دودویی به تقسیمات فرعی و کوچک‌تر تقسیم‌بندی می‌شوند تا زمانی که تقسیمات فرعی نهایی دیگر میسر نباشد. الگوریتم درخت تصمیم، یک

حسب لگاریتم می‌باشد، به دست می‌آید (Rasmussen and William, 2006).

$$\log p(y|X) = -\frac{1}{2} y^T (K + \sigma^2 I)^{-1} y - \frac{1}{2} \log |K + \sigma^2 I| - \frac{n}{2} \log(2\pi) \quad (14)$$

۲-۲-۴- روش یادگیری گروهی

روش یادگیری گروهی شامل ترکیب اطلاعات رگرسیون ۱۰۰ درخت تصمیم (مدل‌های پایه ضعیف) با کمک boosting است. عملیات ترکیب boosting به این صورت است که مدل‌های ضعیف به صورت تکراری و با روش حداقل مربعات آموزش ببینند به طوری که آموزش مدل در یک مرحله، وابسته به مدل‌های آموزش دیده در مراحل قبلی است (Jerome, 2001; Seifi et al., 2022).

۲-۲-۳- معیارهای ارزیابی دقت مدل‌سازی

در پژوهش حاضر، آماره‌های میانگین خطای مطلق (MAE)^{۲۱}، مجذور میانگین مربعات خطا (RMSE)^{۲۲}، ضریب همبستگی (CC)^{۲۳} و ضریب تعیین (R²)^{۲۴} برای ارزیابی نتایج و دقت مدل استفاده شد. به منظور آموزش مدل‌های هوشمند پیش‌بینی میزان تبخیر، ابتدا برای توصیف رفتار دینامیکی پدیده تبخیر از بازسازی فضای فاز استفاده شد. از این رو، با استفاده از الگوریتم نزدیکترین همسایه نادرست، بعد بهینه و با کمک تابع اطلاعات متقابل، تأخیر زمانی بهینه، تخمین زده شد. در نهایت، با استفاده از مقادیر بعد تعبیه و زمان تأخیر، فضای فاز بازسازی شده و از این اطلاعات برای پیش‌بینی مقادیر تبخیر توسط مدل‌های ماشین بردار پشتیبان، درخت تصمیم، رگرسیون فرآیند گاوسی و روش ترکیبی استفاده شد.

ساختار درختی از داده‌های آموزشی ایجاد می‌کند که مجموعه‌ای از تصمیمات را نشان می‌دهد که منجر به پیش‌بینی متغیر هدف می‌شود. این الگوریتم با تقسیم بازگشتی داده‌ها به زیرمجموعه‌ها بر اساس ویژگی‌هایی کار می‌کند که بیشترین اطلاعات را تا زمانی که پی‌بینی یا تصمیم نهایی گرفته شود، فراهم می‌کند. درختان تصمیم مجموعه‌ای از داده‌های معلوم را می‌گیرند و یک درخت تصمیم را از آن استنتاج می‌کنند (Alberg et al., 2012).

$$SDR = sd(T) - \sum \frac{|T_i|}{T} sd(T_i) \quad (9)$$

÷

روش‌های رگرسیون به‌عنوان یکی از روش‌های نوین بهینه‌سازی مدل‌های غیر خطی مورد استفاده بوده و کاربردهای فراوانی دارد (Vapnik, 1995). در روش رگرسیون تغییرات متغیر وابسته بر اساس اطلاعات متغیرهای مستقل دیگر بررسی می‌شود. در مدل‌های رگرسیون فرآیند گاوسی بر این فرض هستند که مشاهدات تنظیم باید حامل اطلاعاتی درباره هم‌دیگر باشد (Pal and Diswal, 2010).

$$y = f(x(k)) + \xi \quad (10)$$

که در رابطه فوق، x و y به ترتیب نشان‌دهنده دامنه‌ای از ورودی‌ها و خروجی‌ها، f تابع وابسته مجهول و ξ مقدار نویز گاوسی با واریانس σ^2 است. پس از تعیین میانگین و کواریانس، پیش‌بینی متغیرهای خروجی برای نمونه از متغیرهای ورودی x^* به صورت توزیع گاوسی پیش‌بینی شده $P(y_*|X, y, x_*)$ با میانگین و تابع کواریانس به شکل زیر ارائه می‌شود.

$$y_* = m(x_*) + K_*^T (K + \sigma_n^2 I)^{-1} (y - m(x_*)) \quad (12)$$

$$\sigma_{y_*}^2 = k_* + -(K + \sigma_n^2 I)^{-1} K_* \quad (13)$$

در روابط فوق، K ماتریس کواریانس و I ماتریس همانی هستند. مقادیر پارامترها معمولاً با پیشینه‌سازی $\log p(y|X)$ که درست‌نمایی بر

Table 2- Input values to predictive models according to phase space reconstruction

جدول ۲- مقادیر ورودی به مدل‌های پیش‌بینی با توجه به بازسازی فضای فاز

	Input				output
	1	2	3	m dimation	
Training data	X_t	X_{t+1}	$X_{t-\tau}$	$X_{t-(m-1)\tau}$	X_{t+1}
	X_{t+1}	X_{t+2}	X_t	$X_{t-(m-2)\tau}$	X_{t+2}

Test data

۳- نتایج

۳-۱- محاسبه نزدیکترین همسایه نادرست و زمان تأخیر پارامتر تبخیر

به منظور بررسی آشوبناکی و ایجاد فضای فاز، مقدار نزدیکترین همسایه نادرست و اطلاعات متقابل برای سری زمانی تبخیر، به ازای مقادیر مختلف تأخیر و بعد، در شکل ۳ و ۴ آورده شده است. با توجه به شکل‌های مذکور، بهترین مقدار تأخیر ۹ و بعد بهینه برای بازسازی فضای فاز ۱۸ تعیین شد.

بستر جذب برای فضای بازسازی شده با بعد ۱۸ و تأخیر ۹ بر روی فضای سه بعدی سری زمانی تبخیر در شکل ۵ ترسیم شده است.

با توجه به مسیر رفت و برگشتی در شکل ۴ و هندسه بستر جذب که شامل قبض و بسطها در محدوده دینامیکی تغییرات تبخیر است و به علاوه، رفتار تقریباً سیكلیک بودن بستر جذب، می‌توان به رفتار شبه‌پریودیک بودن سری زمانی تبخیر پی برد که با تغییرات سری زمانی تبخیر مطابقت دارد. علاوه بر این، رفتار شبه‌پریودیک در نمودار بازگشتی (شکل ۵-چپ) نیز مشهود است. در این نمودار، علاوه بر وجود الگوهای منظم موازی با قطر اصلی، رفتارهای تصادفی نیز، به صورت کیفی مشاهده می‌شود که می‌تواند بازگوکننده رفتار شبه‌پریودیک سری زمانی تبخیر باشد. مقادیر کمی میزان نرخ بازگشت، قطعیت، آنتروپی و LMAX در جدول ۳ آورده شده است.

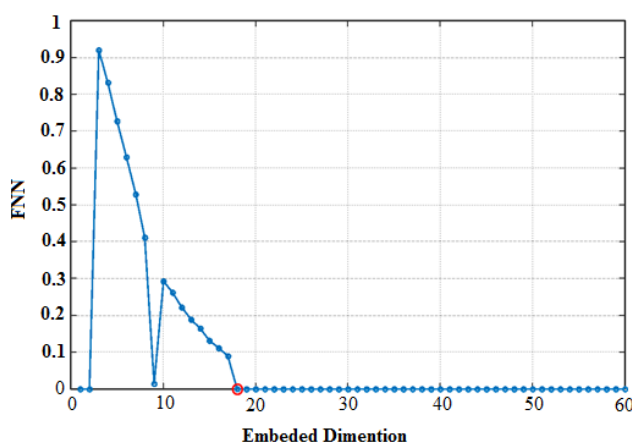


Fig. 3- Variations of false nearest neighbor values for evaporation time series

شکل ۳- تغییرات مقادیر نزدیکترین همسایه نادرست برای سری زمانی تبخیر

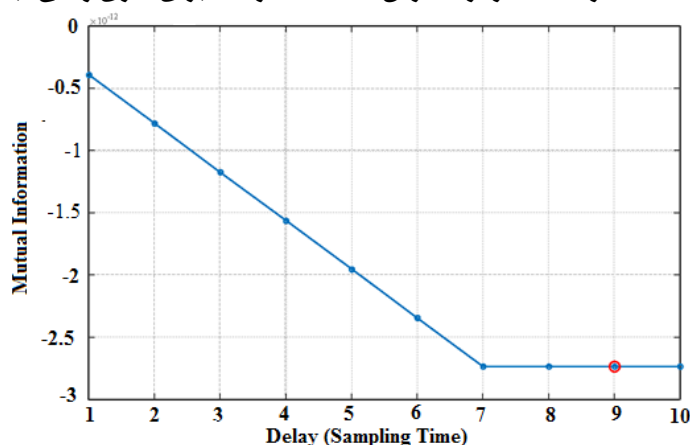
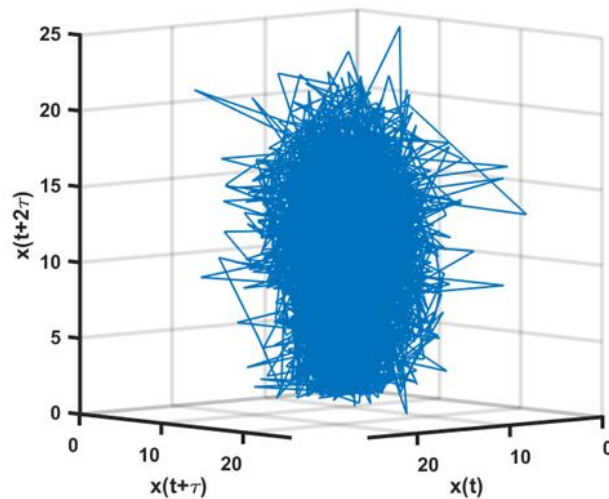


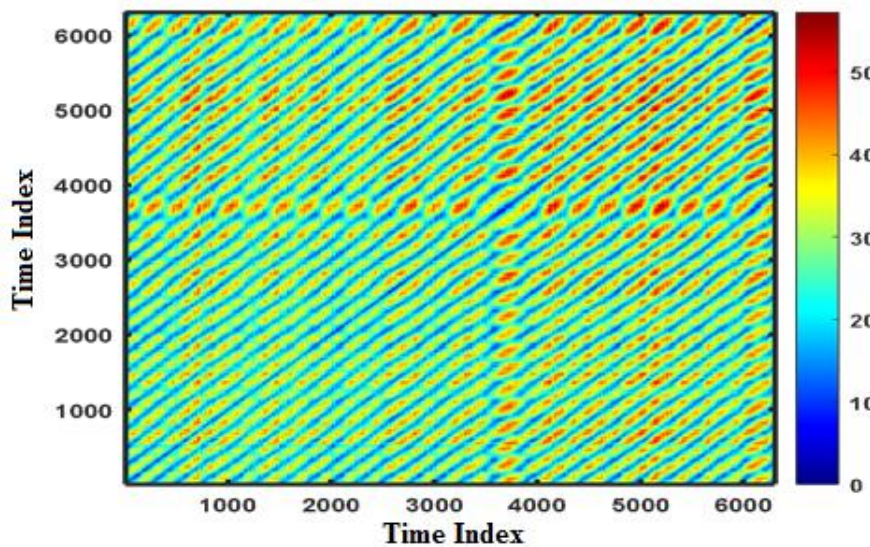
Fig. 4- Mutual information changes for evaporation time series

شکل ۴- تغییرات اطلاعات متقابل برای سری زمانی تبخیر



(a)

Recurrent Plot



(b)

Fig. 5- Absorption bed in the three-dimensional space of the evaporation time series phase space (a) and its recurrence plot (b)

شکل ۵- بستر جذب در فضای سه بعدی فضای فاز سری زمانی تبخیر (a) و نمودار بازگشتی آن (b)

Table 3- Values of return rate, certainty, entropy and LMAX

جدول ۳- مقادیر نرخ بازگشت، قطعیت، آنترپی و LMAX

Time series	Dimension	Delay	LMAX	ENTR	DET	RR
Evaporation	18	9	3.76	1.08	0.75	0.09

دادگان آزمون در جدول ۴ آورده شده است. مقادیر هایپرپارامترهای مرتبط با کرنل گوسین در روش SVM به صورت خودکار با کمک تابع fitsvm در نرم افزار MATLAB انجام شد. نمونه‌ای از میزان

در این قسمت، تنها از دادگان تبخیر زمان‌های قبل، برای صحت‌سنجی استفاده شد. از این‌رو، X یک بردار تک متغیره و تنها شامل دادگان تبخیر بود. نتایج کمی معیارهای کارایی پیش‌بینی میزان تبخیر برای

خواهد بود. با توجه به شکل ۷، میزان تجمع دادگان، تقریباً حول قطر اصلی است که نشان‌دهنده کیفیت بالای مدل پیش‌بینی‌کننده با ضریب تعیین ۸۲/۱۲٪ است.

پیش‌بینی در طی ۶۰ روز در شکل ۶ آورده شده است. نمودار پراکندگی میزان تبخیر پیش‌بینی با مقادیر اندازه‌گیری شده در شکل ۷ آورده شده است. در نمودار پراکندگی، هر میزان داده‌های پیش‌بینی شده با مقادیر اندازه‌گیری شباهت داشته باشند، پراکندگی دادگان حول قطر اصلی

Table 4- Efficiency criteria for predicting evaporation

جدول ۴- معیارهای کارایی پیش‌بینی میزان تبخیر

Data	MAE (mm)	RMSE (mm)	CC (%)	R ² (%)
Training data	1.81	2.45	85.94	73.69
Test data	1.64	2.16	90.66	82.12

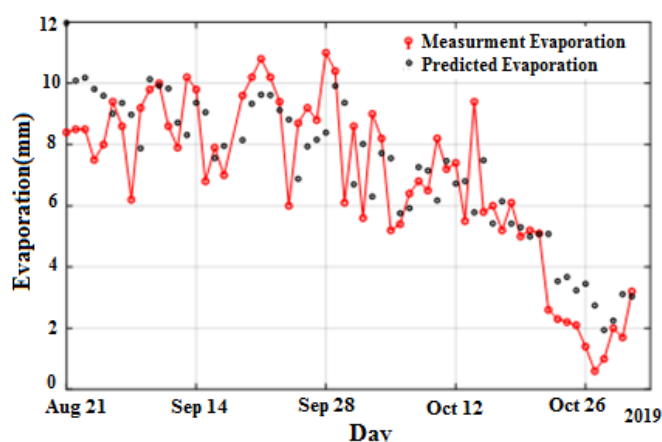


Fig. 6- prediction of evaporation during 60 days using evaporation time series

شکل ۶- پیش‌بینی تبخیر در طی ۶۰ روز با سری زمانی تبخیر

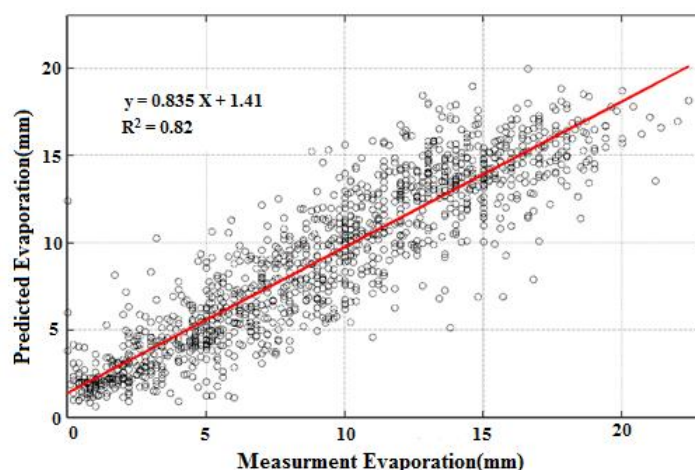


Fig. 7- Scatter plot of predicted vs measured evaporation

شکل ۷- نمودار پراکندگی میزان تبخیر پیش‌بینی شده به اندازه‌گیری شده با ورودی تبخیر

۳-۲- آموزش مدل‌های رگرسیون با ورودی‌هایی با ابعاد بالاتر

از آنجا که در قسمت قبلی تنها از دادگان تبخیر استفاده شده بود، در ادامه از ترکیب متغیرهای دیگری غیر از تبخیر، مانند دما، سرعت باد، فشار، بارش، رطوبت نسبی و ساعات آفتابی به منظور پیش‌بینی میزان تبخیر بهره برده شد. به منظور تخمین پارامترهای فضای فاز برای سری زمانی‌های چندبعدی (MTS)^{۲۵} از روش والت استفاده شد (Walt et al., 2018). نحوه نام‌گذاری ترکیب متغیرهای مختلف به همراه مقادیر بهینه میزان تأخیر و بعد برای بازسازی فضای فاز، در جدول ۵ آورده شده است. با توجه به نتایج، هر میزان که تعداد متغیرهای در نظر گرفته افزایش پیدا می‌کنند، میزان بعد و تأخیر لازم برای بازسازی فضای فاز کاهش پیدا می‌کند.

از سه مدل رگرسیونی برای صحت‌سنجی مبتنی بر SVM استفاده شد. رگرسورها شامل درخت تصمیم (DT)، رگرسیون ترکیبی مبتنی بر یادگیری گروهی (EL) و رگرسیون فرآیند گوسی (GP) بودند. نتایج معیارهای کمی میزان پیش‌بینی برای دادگان آزمون توسط روش‌های مختلف در جدول ۶ آورده شده است. در اجرای الگوریتم‌های مورد بررسی، ترکیب‌های متفاوتی از متغیرهای مستقل مورد ارزیابی قرار گرفتند. بهینه‌ترین پاسخ برای ترکیب کلیه پارامترها شامل (تبخیر، دما، سرعت باد، فشار، سطح بارندگی، رطوبت و ساعات آفتابی) با داشتن ضرایب عملکرد $R^2 = 85.58$, $CC = 92.67$, $RMSE = 1.95$ و $MAE = 1.46$ در مدل SVR حاصل شد. منحنی پراکندگی میزان تبخیر پیش‌بینی‌شده به اندازه‌گیری شده با توجه به اطلاعات ترکیب‌شده سری زمانی‌های مختلف برای مدل SVR در شکل ۷ نشان داده شده است.

در روش درخت تصمیم (DT) نیز ترکیب مورد استفاده، منجر به بهترین انطباق با آماره‌های اندازه‌گیری و بدست آمدن مقادیری برابر با $R^2 = 83.07$, $CC = 91.18$, $RMSE = 2.1$ و $MAE = 1.62$ شد. نتایج روش یادگیری گروهی (EL) نیز مانند دو مدل SVR و DT برای مناسبترین ترکیب پارامترها که شامل همه آنها بود ضرایب عملکرد $R^2 = 85.37$, $CC = 92.57$, $RMSE = 1.96$ و $MAE = 1.48$ را نشان داد. بر اساس نتایج بدست آمده، در روش رگرسیون فرآیند گوسی (GP) ترکیب همه‌ی پارامترها با داشتن ضرایب عملکرد $R^2 = 85.46$, $CC = 92.56$, $RMSE = 1.96$ و $MAE = 1.48$ بهترین ترکیب شناخته شد. لازم به ذکر است که ترکیب پارامترهای تبخیر و دما در مقایسه با استفاده از همه‌ی پارامترهای هواشناسی شامل تبخیر، دما، سرعت باد، فشار، سطح بارندگی، رطوبت و ساعات آفتابی برای تمامی مدل‌ها عملکرد مناسب و قابل قبولی نشان داد و اختلاف زیادی نداشت. به طور کلی، با توجه به نتایج، هر چند عملکرد کلیه رگرسورها مناسب می‌باشد ولی مدل SVR توانست پیش‌بینی مناسب‌تری با ضرایب قطعیت و همبستگی بیشتر و در همین حال مقدار کمتر $RMSE$ و MAE را مهیا سازد. در مجموع می‌توان گفت اگر از متغیرهای بیشتر به همراه تأخیرهای مناسب آنها به عنوان مدل ورودی پیش‌بینی تبخیر استفاده شود، می‌توان مدل جامع‌تری برای پیش‌بینی ارائه داد، که با استفاده از اطلاعات بازسازی فضای فاز، که تأخیر بهینه را فراهم می‌سازد، این امر مهیا شده است.

نمودار پراکندگی میزان تبخیر پیش‌بینی شده با مقادیر اندازه‌گیری شده در شکل ۸ آورده شده است. بر طبق نتایج مدل‌های مختلف استفاده شده برای صحت‌سنجی، نتایج مدل SVM که عملکرد بهترین عملکرد را داشت نشان‌دهنده بیشترین پراکنش داده‌ها حول قطر اصلی یا خط برازش با همبستگی بالا ملاحظه شد.

Table 5- Delay and dimension parameters to reconstruct the state space for different MTSs

جدول ۵- پارامترهای تأخیر و بعد به منظور بازسازی فضای حالت برای MTS های مختلف

Scenario (multidimensional time series)	Dimension	Delay
Evaporation/Temperature	MTS1	5
Evaporation/Temperature/Wind Speed	MTS2	3
Evaporation/Temperature/Wind Speed/Pressure	MTS3	2
Evaporation/temperature/wind speed/pressure/precipitation	MTS4	2
Evaporation/Temperature/Wind Speed/Pressure/Precipitation	MTS5	2
/Humidity		
Evaporation/Temperature/Wind Speed/Pressure/Rain	MTS6	2
/Humidity/Sunshine Hours		

Table 6- Results of quantitative criteria for four regressors (SVR, DT, EL and GP)

جدول ۶- نتایج معیارهای کمی برای ۵ رگرسور SVR، DT، EL و GP

Model / Regressor	GP				EL				DT				SVR			
	MAE (mm)	RMSE (mm)	CC (%)	R2 (%)	MAE (mm)	RMSE (mm)	CC (%)	R2 (%)	MAE (mm)	RMSE (mm)	CC (%)	R2 (%)	MAE (mm)	RMSE (mm)	CC (%)	R2 (%)
MTS1	1.59	2.08	91.5	83.46	1.6	2.07	91.62	83.54	1.65	2.15	90.82	82.31	1.54	2.04	91.74	84.02
MTS2	1.55	2.05	91.71	83.93	1.55	2.05	91.83	83.97	1.68	2.21	90.25	81.35	1.53	2.04	91.77	84.17
MTS3	1.54	2.04	91.97	84.13	1.58	2.06	91.79	83.81	1.67	2.19	90.49	81.76	1.52	2.03	92.07	84.17
MTS4	1.53	2.03	92.02	84.24	1.52	2.01	92.18	84.61	1.66	2.19	90.49	81.7	1.52	2.05	92.04	84.01
MTS5	1.52	2.01	92.17	84.69	1.55	2.01	92.14	84.51	1.66	2.16	90.71	82.14	1.5	2.01	92.27	84.8
MTS6	1.48	1.96	92.56	85.46	1.48	1.96	92.59	85.37	1.62	2.1	91.18	83.07	1.46	1.95	92.67	85.58

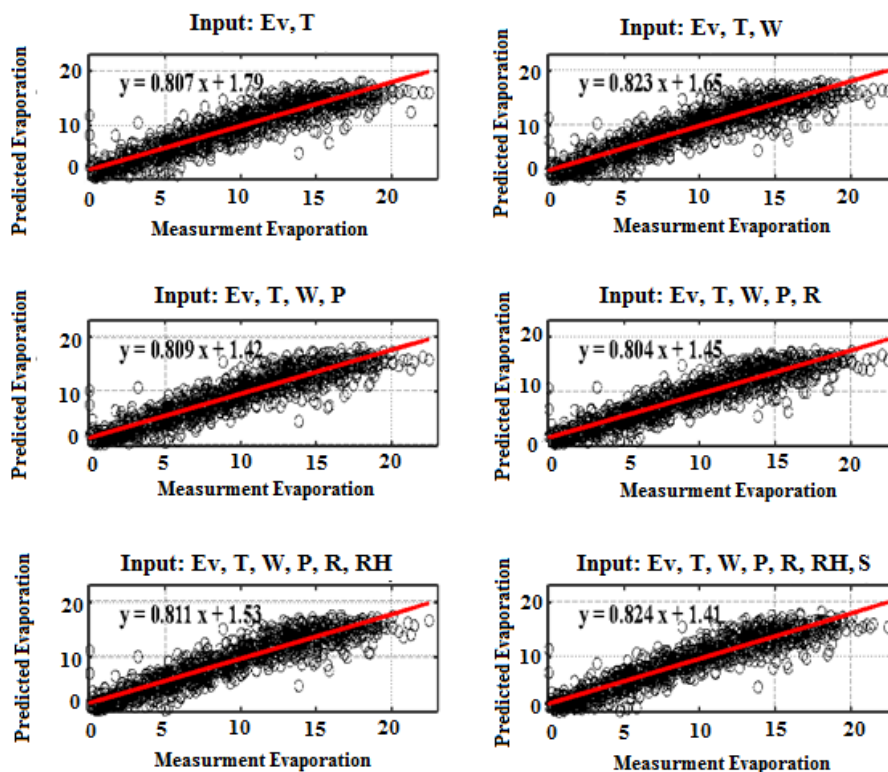


Fig. 8- Scatter plot of predicted vs measured evaporation with hybrid data of different time series for SVR (Ev: Evaporation, T: Temperature, W: Wind speed, P: Pressure, R: Rainfall, RH: Relative Humidity, S: Sunshine Hours)

شکل ۸ - نمودار پراکنندگی میزان تبخیر پیش‌بینی شده به اندازه‌گیری شده با توجه به اطلاعات ترکیب شده سری زمانی‌های مختلف SVR

(Ev: تبخیر، T: دما، W: سرعت تبخیر، P: فشار، R: بارش، RH: رطوبت نسبی، S: ساعات آفتابی)

اهمیت زیادی دارد و این تحقیق با هدف شناخت بهتر ماهیت دینامیک پدیده تبخیر و پیش‌بینی آن با استفاده از تئوری آشوب در ترکیب با مدل‌های هوشمند انجام شد.

در پژوهش حاضر، تلفات آب از طریق تبخیر با استفاده از مدل‌های SVR، DT، EL و GP پیش‌بینی شد. داده‌های ورودی برای مدل‌ها با استفاده از یک رویکرد جدید و متناسب با ماهیت پدیده تبخیر، یعنی نظریه آشوب تهیه شد. برای استفاده از مزایای نظریه آشوب، لازم است فضای فاز با استفاده از مشاهدات بازسازی شود. بنابراین، در این مطالعه، دو نوع تکنیک تعبیه فضای فاز برای مشاهدات اعمال شد. این تکنیک‌ها تعبیه سری‌های زمانی تک متغیره و چند متغیره هستند. مهمترین نتایج بدست آمده عبارتند از:

- با بررسی نزدیکترین همسایه نادرست و زمان تأخیر، مقدار تأخیر ۹ و بعد بهینه برای فضای بازسازی ۱۸ تعیین شد؛

نحوه صحت‌سنجی با توجه به اطلاعات ترکیب‌شده سری زمانی‌های مختلف برای مدل SVR در شکل ۹ مشخص است. بر طبق نتایج در این مطالعه، منحنی‌های مربوط به مقادیر اندازه‌گیری شده و پیش‌بینی شده انطباق خوبی با هم دارند و اختلاف کمی بین آنها مشاهده می‌شود که حاکی از دقت بالای مدل در پیش‌بینی است. با دقت در این نمودارها مشخص می‌شود که نتایج مدل SVR برای نقاط بیشینه و کمینه کمتر از مقادیر اندازه‌گیری شده هستند.

۴- جمع‌بندی

تبخیر یکی از مهم‌ترین اجزای تشکیل‌دهنده تلفات آب است که تابع عوامل مختلف بوده و به علت تأثیرپذیری به متغیرهای آب و هوایی متفاوت و اثر متقابل این متغیرها بر همدیگر، یک پدیده غیر خطی و پیچیده است و در مطالعه آن باید روش‌های دقیقی را برای شبیه‌سازی به کار گرفت. پیش‌بینی تبخیر در مدیریت و برنامه‌ریزی اصولی آب

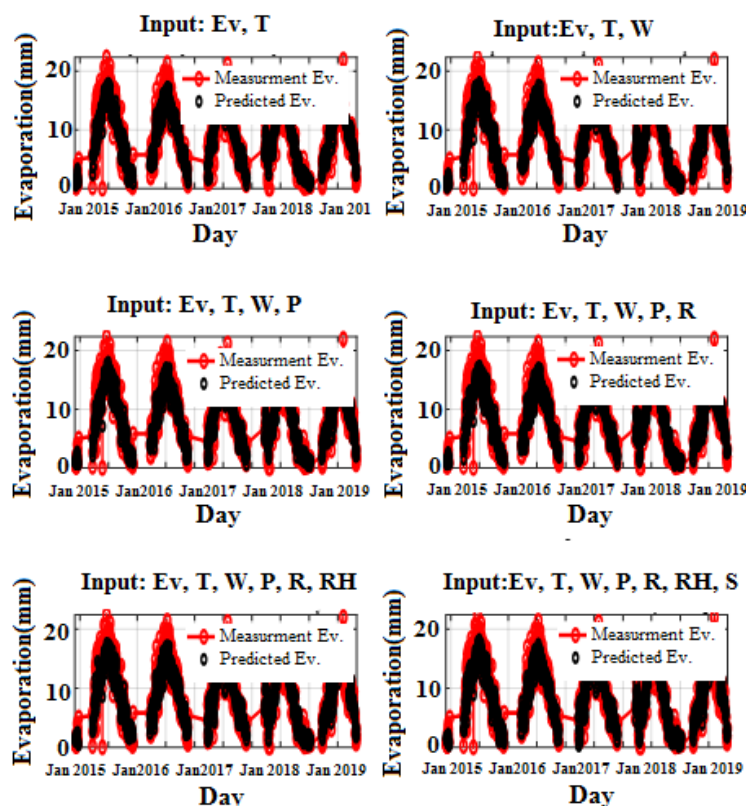


Fig. 9- Prediction of evaporation during 1292 days with the combined data time series using SVR (Ev: Evaporation, T: Temperature, W: Wind speed, P: Pressure, R: Rainfall, RH: Relative Humidity, S: Sunshine Hours)

شکل ۹- پیش‌بینی میزان تبخیر طی ۱۲۹۲ روز با اطلاعات ترکیب‌شده سری زمانی‌های مختلف SVR (Ev: تبخیر، T: دما، W: سرعت تبخیر، P: فشار، R: بارش، RH: رطوبت نسبی، S: ساعات آفتابی)

می‌تواند در برنامه‌ریزی مناسب‌تر منابع آب در کشور توسط مدیران این بخش سودمند باشد.

پی‌نوشت‌ها

- 1- Support Vector Machine (SVM)
- 2- Chaos Theory
- 3- Multivariate Regression Spline (MRS)
- 4- Deep Learning (DL)
- 5- Random Forest (RF)
- 6- Long Short-Term Memory (LSTM)
- 7- Multilayer Perceptron (MLP)
- 8- Seasonal Autoregressive Integrated Moving Average (SARIMA)
- 9- Copula Bayesian Model (CBM)
- 10- Machine Learning (ML)
- 11- Decision Tree (DT)
- 12- Gaussian Process (GP)
- 13- Ensemble of Learners (EL)
- 14- Phase Space (PS)
- 15- False Nearest Neighbor (FNN)
- 16- Recurrent Quantification Analysis (RQA)
- 17- Recurrence Rate (RR)
- 18- Entropy
- 19- Maximum Line (LMAX)
- 20- Positive Lyapunov Exponent (PLE)
- 21- Mean Absolute Error (MAE)
- 22- Root Mean Square Error (RMSE)
- 23- Correlation Coefficient (CC)
- 24- Determinant Coefficient (DC)
- 25- Multidimensional Time Series (MTS)

- ترکیب‌های متفاوتی از متغیرهای مستقل مورد ارزیابی قرار گرفتند که بهینه‌ترین پاسخ برای ترکیب کلیه پارامترها (شامل تبخیر، دما، سرعت باد، فشار، ارتفاع بارندگی، رطوبت و ساعات آفتابی) با داشتن ضرایب عملکرد $R^2 = 85.58$, $CC = 92.67$, $RMSE = 1.95$ و $MAE = 1.46$ در مدل SVR تعیین شد؛

- مدل SVR توانست دقت مناسب‌تری با ضرایب قطعیت و همبستگی بیشتر و در همین حال مقدار کمتر RMSE و MAE مهیا سازد که با نتایج Ozlem and Kassim (2014) که به پیش‌بینی تبخیر بر اساس SVR همراه با رویکرد آشوب پرداختند مطابقت دارد. همچنین با تحقیقات Wang et al. (2017) که یک روش ترکیبی بر اساس طرح تکرار و رگرسیون بردار پشتیبان بهینه (SVR) برای پیش‌بینی سرعت باد استفاده کردند و همچنین مطالعات El Bilali et al. (2023) و Wang et al. (2007) مطابقت دارد.

با توجه به اینکه در تمامی روش‌ها میزان تأخیر برای تمام رگرورها لحاظ شده است و پارامترهای هر روش به صورت بهینه‌ای تخمین زده شده است، نتایج روش‌های مختلف نتایج تقریباً مشابهی را موجب شده‌اند. لازم به ذکر است با توجه به اینکه کاهش نویز در این مطالعه انجام نشده است، اما ممکن است در صورتی که مدل‌های مورد بررسی به خصوص مدل SVR برای داده‌های بدون نویز اعمال شود پیش‌بینی‌هایی با دقت بالاتر حاصل شود. لذا انجام عملیات پیش‌بینی با استفاده از داده‌های نویزگیری شده پیشنهاد می‌شود. همچنین با توجه به اثرات محتمل تغییر اقلیم بر چرخه آب خصوصاً میزان تبخیر از روش‌های مبتنی بر یادگیری عمیق مانند شبکه‌های عصبی حافظه طولانی کوتاه مدت (LSTM) برای پیش‌بینی میزان تأخیر می‌توان بهره برد. یافته‌های حاصل از پژوهش حاضر، با توجه به دقت مناسب روش‌های هیبریدی توسعه یافته که مورد استفاده قرار گرفت، برای سایر مناطق به منظور پیش‌بینی تبخیر برای آینده قابل استفاده می‌باشد. با توجه به مسئله تبخیر و کمبود آب در کشور، پیش‌بینی پدیده تبخیر و برنامه‌ریزی متناسب جهت کاهش تبخیر و مدیریت آن در سال‌های آتی ضروری است. لذا، استفاده از نتایج بدست آمده

۵- مراجع

- Alberg D, Last M, Kandel A (2012) Knowledge discovery in data streams with regression tree methods. *WIREs Data Mining Know Discover* 2:69-78
- Alsumaiei A A (2020) Utility of artificial neural networks in modeling pan evaporation in hyper-arid climates. *Water* 12(5):1508
- Ayodele T R, Ogunjuyigbe A S O, Amedu A, Munda J L (2019) Prediction of global solar irradiation using hybridized k-means and support vector regression algorithms. *Renwable Energy Focus* 29:78-93
- Benzaghta M A (2014) Estimation of evaporation from a reservoir in semi arid environments using artificial neural network and climate based models. *Journal of Applied Science and Technology* 4(24):3501-3518
- El Bilali A, Abdeslam A, Ayoub N, Lamane H, Abdellah Ezzaouini M, Elbeltagi A (2023) An interpretable machine learning approach based on DNN, SVR, Extra Tree, and XGBoost models for predicting daily pan evaporation. *Journal of Environmental Management* 327:116890
- Farzin S, Farzin N, Hassanzadeh Y, Aalami MT (2012) An investigation on changes and prediction of Urmia Lake water surface evaporation by Chaos Theory. *International Journal of Environmental Reaserch* 6(3):815-824
- Farzin S, Hajiabadi, R, Ahmadi M H (2017) Application of Chaos Theory and artificial neural networks to evaluate evaporation from lake's water surface. *Water and Soil* 31(1):61-74
- Ferreira L B, Da Cunha F, OliveraD R, Fernandes EI (2019) Estimation of reference evapotranspiration in Brazil with limited meteorological data using ANN and SVM–A new approach. *Journal of Hydrology* 527:556-570
- Jerome H (2001) 1999 Reitz lecture greedy function approximation: a gradient boosting machine. *The Annals of Statistis* 29 (5):1189-1232
- Kisi O (2015) Pan evaporation modeling using least square support vector machine multivariate adaptive regression splines and M5 model tree. *Journal of Hydrology* 528:312-320
- Kumar Goyal M, Birendra B, Quilty J, Adamowksi J, Pandey A (2017) Modeling of daily pan evaporation in sub tropical climates using ANN, LS-SVR, Fuzzy Logic, and ANFIS. *Expert Systems with Applications* 41:5267-5276
- Kocak K, Bali A, Bektasoglu B (2007) Prediction of monthly flows by using chaotic approach. *International Congress on River Basin Management, 22-24 March, Antalya, Turkey, p. 553-559*
- Khullar S, Singh N (2022) Water quality assessment of a river using deep learning Bi-LSTM methodology: Forecasting and validation. *Environmental Science and Pollution Research* 29(9):12875-89
- La Fuente S, Jennings E, Gal G, Kirillin G, Shatwell T, Ladwig R, Moore T, Couture T. M. Cote M, Vinna R, Woolway I (2022) Multi-model projections of future evaporation in a sub-tropical lake. *Journal of Hydrology* 615:Part A, 128729
- Malik A, Kuma, A, Kisi O (2019) Monthly pan-evaporation estimation in Indian central Himalayas using different heuristic approaches and climate based models. *Computers and Electronics in Agriculture* 143:302-313
- Ozlem B, Kasim K (2014) SVR-based prediction of evaporation combined with chaotic approach. *Journal of Hydrology* 508:356-363
- Pal M, Deswal S (2010) Modelling pile capacity using Gaussian process regression. *Computers and Geotechnics* 37:942-947
- Rasmussen C E, Williams C K (2006) *Gaussian process for machine learning*. Massachusetts. MIT Press, p 266.
- Rocha S, Molinas E, Rodrigues I (2023) Assessment of total evaporation rates and its surface distribution by tridimensional modelling and remote sensing. *Journal of Environmental Management* 327:116846
- Sik Kim M, Cha D, Mi Lee S, Jeong H, Lee L (2023) Prediction of brine evaporation rate in a pond: Development of different models under controlled meteorological conditions and comparative evaluation. *Desalination* 551:116415
- Taghipour K (2023) Application of chaos theory in electrical engineering. *Journal of Science Engineering Elites* 8(1):35-48
- Wang Y, Jiang T, Bothe O, Fraedrich K (2007) Changes of pan evaporation and reference evapotranspiration in the Yangtze River basin. *Theoretical and Applied Climatology* 90(1-2):13-23
- Wang W, Xu D, Chau K, Chen S (2013) Improved annual rainfall-runoff forecasting using PSO-SVM model based on EEMD. *Journal of Hydroinformatics* 15:1377–1390
- Webber C L, Zbilut J P (2005) Recurrence quantification analysis of nonlinear dynamical systems. *Tutorials in Contemporary Nonlinear Methods for the Behavioral Sciences* 26–94

- Wang L, Niu Z, Kisi O, Li C, Yu D (2017) Pan evaporation modeling using four different heuristic approaches. *Computers and Electronics in Agriculture* 140:203-213
- Vapnik V (1995) *The nature of statistical learning theory*. Data Mining and Knowledge Discovery, Springer, Berlin
- Wallot S, Mønster D (2018) Calculation of Average Mutual Information (AMI) and False-Nearest Neighbors (FNN) for the estimation of embedding parameters of multidimensional time series in matlab. *Frontiers in psychology* 10(9): 1679
- hang H, Gorelick S M, Zimba PV, Zhang X (2017) A remote sensing method for estimating regional reservoir area and evaporative loss. *Journal of Hydrology* 555:213-227
- Zhao G, Gao H (2019) Estimating reservoir evaporation losses for the United States: Fusing remote sensing and modeling approaches. *Remote Sensing of Environment* 226:09-124