

کاربرد بازسازی فضای فاز در بهبود کارایی مدل‌های یادگیری

ماشینی در پیش‌بینی خشکسالی

(منطقه مورد مطالعه: ایستگاه سینوپتیک بندرعباس)

(مقاله پژوهشی)

۱- مهرناز یحیی‌زاده، دانشجوی دکتری گروه مهندسی منابع طبیعی، دانشکده کشاورزی و منابع طبیعی، دانشگاه هرمزگان، بندرعباس، ایران.

۲- ام‌البنین بذرافشان*، استاد گروه مهندسی منابع طبیعی، دانشکده کشاورزی و منابع طبیعی، دانشگاه هرمزگان، بندرعباس، ایران.

O.bazrafshan@hormozgan.ac.ir

۳- نوازاله مرادی، استادیار گروه مهندسی منابع طبیعی، دانشکده کشاورزی و منابع طبیعی، دانشگاه هرمزگان، بندرعباس، ایران.

۴- حسین زمانی، استادیار، گروه ریاضی و آمار، دانشکده علوم پایه، دانشگاه هرمزگان، بندرعباس، ایران.

۵- مهدی بی‌نیاز، استادیار گروه مهندسی منابع طبیعی، دانشکده کشاورزی و منابع طبیعی، دانشگاه هرمزگان، بندرعباس، ایران.

دریافت: ۱۴۰۴/۰۹/۰۸

پذیرش: ۱۴۰۴/۱۱/۱۲

چکیده

خشکسالی به‌عنوان یکی از پدیده‌های اقلیمی پیچیده و چندبعدی، همواره با عدم قطعیت بالا و وابستگی‌های غیرخطی میان متغیرهای اقلیمی همراه است. در این پژوهش، به‌منظور بهبود دقت و تفسیرپذیری پیش‌بینی خشکسالی در اقلیم فراهشک بندرعباس طی دوره آماری ۱۹۶۰ تا ۲۰۲۲ با استفاده از چارچوب ترکیبی متشکل از بازسازی فضای فاز (PSR)، واین‌کاپولا (Vine Copula) و رگرسیون چندکی توسعه داده شد. نتایج نشان داد که بازسازی فضای فاز با آشکارسازی ساختار دینامیکی پنهان سری زمانی شاخص SPEI موجب بهبود چشمگیر عملکرد مدل‌ها شد، به‌گونه‌ای که مقدار R^2 در مرحله آزمون از حدود ۰/۱ به بیش از ۰/۸۰ افزایش یافت. تحلیل واین‌کاپولا وابستگی‌های غیرخطی و نامتقارن میان متغیرهای اقلیمی را به‌طور مؤثر مدل‌سازی کرد و در ترکیب با PSR، چارچوبی منسجم برای بازنمایی رفتار پویای سامانه خشکسالی فراهم آورد. نسخه‌های مبتنی بر رگرسیون چندکی شامل QRF، QXGBoost و QVineCopula، افزون بر بهبود دقت پیش‌بینی، امکان کمی‌سازی داده‌محور عدم قطعیت را فراهم کرده و پوشش بهتر در بازه‌های پیش‌بینی را نشان دادند ($PICP = 0.91$). تحلیل اهمیت متغیرها بیانگر نقش غالب مؤلفه‌های تاخیری SPEI، به‌ویژه SPEI9 و SPEI8، در کنار دمای بیشینه و سرعت باد، در پویایی خشکسالی منطقه بود. در مجموع، مدل ترکیبی PSR-QVineCopula تعادلی بهینه میان دقت، پایداری و تفسیرپذیری برقرار نموده و چارچوبی نوین برای پیش‌بینی‌های قطعی و احتمالی خشکسالی در مناطق فراهشک ارائه می‌دهد.

واژگان کلیدی: خشکسالی، بازسازی فضای فاز (PSR)، واین‌کاپولا، رگرسیون چندکی، پیش‌بینی احتمالی، عدم قطعیت.

مقدمه

خشکسالی یکی از پیچیده‌ترین، فراگیرترین و پرهزینه‌ترین پدیده‌های هیدرو-اقلیمی است که به‌طور مستقیم حیات انسان، پایداری اکوسیستم‌های طبیعی و امنیت غذایی را تحت تأثیر قرار می‌دهد.

استاندارد شده (SPEI) اشاره کرد. SPI صرفاً بر پایه بارش تعریف شده و شدت خشکسالی را در مقیاس‌های زمانی مختلف (نظیر ۱، ۶ و ۱۲ ماهه) توصیف می‌کند، در حالی که SPEI با در نظر گرفتن تبخیر و تعرق پتانسیل، اثر شرایط دمایی و گرمایش اقلیمی را نیز وارد تحلیل می‌نماید و از این رو برای مناطق گرم و فراخشک، از جمله جنوب ایران، شاخص مناسب‌تری محسوب می‌شود. با این وجود، ماهیت تدریجی، غیرایستا و چندمتغیره خشکسالی سبب شده است که تحلیل‌های تک‌متغیره و خطی، توانایی محدودی در بازنمایی پویایی واقعی این پدیده داشته باشند.

خشکسالی ذاتاً یک پدیده چندمتغیره است و معمولاً میان متغیرهای اقلیمی مؤثر بر آن، همبستگی‌های معنی‌دار و در عین حال غیرخطی وجود دارد. این ویژگی سبب شده است که استفاده از تحلیل‌های آماری چندمتغیره، درک عمیق‌تری از رفتار خشکسالی فراهم آورد. با این حال، از آنجا که توزیع احتمال متغیرهای اقلیمی و شاخص‌های خشکسالی لزوماً یکسان یا هم‌خانواده نیست، استفاده از توزیع‌های چندمتغیره کلاسیک با محدودیت‌های جدی مواجه می‌شود. در این میان، توابع کاپولا به‌عنوان ابزاری انعطاف‌پذیر، امکان مدل‌سازی ساختار وابستگی میان متغیرها را بدون الزام به یکسان بودن توزیع‌های حاشیه‌ای فراهم می‌کنند [۲۹]. ویژگی کلیدی کاپولاها، جداسازی توزیع‌های حاشیه‌ای از ساختار وابستگی بوده که اجازه می‌دهد متغیرهایی با توزیع‌های متفاوت، در چارچوبی واحد و سازگار ترکیب شوند.

در سال‌های اخیر، استفاده از ساختارهای پیشرفته کاپولا، به‌ویژه واین‌کاپولاها شامل C-vine, D-vine و R-vine، به‌طور گسترده در تحلیل وابستگی‌های هیدرو-اقلیمی مورد توجه قرار گرفته است [۹]. این ساختارها با تجزیه وابستگی چندبعدی به مجموعه‌ای از کاپولاها دوتایی شرطی، توانایی بالایی در بازنمایی روابط غیرخطی، نامتقارن و وابستگی‌های

برخلاف مخاطرات ناگهانی، خشکسالی ماهیتی تدریجی، خزنده و چندبعدی دارد و معمولاً با عدم قطعیت بالا، روابط غیرخطی و برهم‌کنش پیچیده میان متغیرهای اقلیمی همراه است. کاهش مقدار و تداوم بارش، افزایش دمای سطحی و تبخیر-تعرق، و تغییر در شدت و الگوی باد، از جمله عوامل اصلی تشدیدکننده خشکسالی به‌شمار می‌روند که در نهایت می‌توانند منجر به افت رطوبت خاک، کاهش عملکرد محصولات کشاورزی و افزایش تنش آبی در سامانه‌های طبیعی و انسانی شوند [۴۲].

از منظر مفهومی، خشکسالی به چهار نوع اصلی شامل خشکسالی هواشناسی، کشاورزی، هیدرولوژیکی و اقتصادی-اجتماعی تقسیم می‌شود که هر یک دارای ویژگی‌ها، سازوکارها و پیامدهای متفاوتی هستند. خشکسالی هواشناسی زمانی رخ می‌دهد که میزان بارش دریافتی یک منطقه در یک بازه زمانی مشخص، به‌طور معنی‌دار کمتر از میانگین بلندمدت باشد [۸، ۱۳]. خشکسالی کشاورزی به کمبود رطوبت در ناحیه ریشه گیاهان اشاره دارد که حتی در شرایطی که بارندگی از نظر آماری نرمال به نظر می‌رسد، می‌تواند رشد و عملکرد محصولات زراعی را محدود سازد [۱۸، ۳۸]. خشکسالی هیدرولوژیکی با کاهش جریان رودخانه‌ها، افت سطح آب‌های زیرزمینی و کاهش ذخایر منابع آب سطحی و زیرسطحی مشخص می‌شود [۱۴، ۲۶]. نهایتاً، خشکسالی اقتصادی-اجتماعی زمانی پدیدار می‌شود که کمبود منابع آب، ناشی از انواع پیشین خشکسالی، بر عرضه محصولات کشاورزی، امنیت غذایی، اشتغال و ساختار اقتصادی جوامع اثر منفی بگذارد [۲۲، ۴۰]. در عمل، این انواع خشکسالی به‌صورت زنجیره‌ای و با تأخیر زمانی نسبت به یکدیگر بروز می‌کنند که پیش‌بینی و مدیریت آن‌ها را به چالشی پیچیده تبدیل می‌نماید.

برای پایش و تحلیل خشکسالی اقلیمی، شاخص‌های متعددی توسعه یافته‌اند که از پرکاربردترین آن‌ها می‌توان به شاخص بارش استاندارد شده (SPI) و شاخص بارش-تبخیر و تعرق

احتمالی فراهم می‌آورد. مطالعات پیشین نشان داده‌اند که مدل‌های مبتنی بر واین‌کاپولا و رگرسیون چندکی، عملکرد برتری نسبت به رویکردهای کلاسیک و حتی بسیاری از مدل‌های یادگیری ماشینی در تحلیل خشکسالی و متغیرهای هیدرو-اقلیمی دارند [۱، ۲۵، ۲۶، ۳۷].

با وجود این پیشرفت‌ها، مرور ادبیات نشان می‌دهد که اغلب مطالعات، یا بر مدل‌سازی وابستگی چندمتغیره تمرکز داشته‌اند، یا بر پیش‌بینی مبتنی بر یادگیری ماشینی، و به‌ندرت چارچوبی منسجم ارائه شده است که به‌طور همزمان دینامیک زمانی سیستم (از طریق PSR)، ساختار وابستگی غیرخطی (از طریق واین‌کاپولا) و عدم قطعیت پیش‌بینی (از طریق رگرسیون چندکی) را در یک چارچوب یکپارچه در نظر گیرد؛ به‌ویژه در اقلیم‌های فراخشک که رفتار خشکسالی غیرخطی‌تر و پیچیده‌تر است.

بر این اساس، هدف مطالعه حاضر توسعه یک چارچوب ترکیبی مبتنی بر بازسازی فضای فاز، واین‌کاپولا و رگرسیون چندکی به‌منظور تحلیل ساختار وابستگی چندمتغیره، بهبود دقت پیش‌بینی قطعی و احتمالی خشکسالی، و افزایش تفسیرپذیری مدل در اقلیم فراخشک بندرعباس است. این چارچوب تلاش می‌کند با اتصال مفاهیم دینامیک سیستم‌ها، آمار چندمتغیره و یادگیری ماشینی، شکاف موجود میان دقت پیش‌بینی و درک فیزیکی رفتار خشکسالی را پر نماید.

داده‌ها و روش‌شناسی

منطقه مورد مطالعه و داده‌ها

شهر بندرعباس مرکز استان هرمزگان و یکی از مهم‌ترین شهرهای ساحلی جنوب ایران است که در حاشیه شمالی تنگه هرمز و در کرانه‌های شمالی خلیج فارس واقع شده است. این شهر به‌صورت نواری ساحلی توسعه یافته و از موقعیت جغرافیایی و اقلیمی منحصربه‌فردی برخوردار است. مختصات ایستگاه سینوپتیک منطقه در حدود ۲۷ درجه عرض شمالی و ۵۶ درجه طول شرقی قرار دارد. مساحت محدوده

دمی دارند؛ ویژگی‌ای که در تحلیل خشکسالی، به‌ویژه در رخدادهای حدی، اهمیت ویژه‌ای دارد.

هم‌زمان با توسعه روش‌های آماری چندمتغیره، مدل‌های یادگیری ماشینی به‌عنوان ابزارهایی قدرتمند و داده‌محور برای مدل‌سازی پدیده‌های پیچیده هیدرو-اقلیمی معرفی شده‌اند. این مدل‌ها، از جمله جنگل تصادفی (RF)، ماشین‌بردار پشتیبان (SVM)، XGBoost و M5Tree، قادرند الگوهای غیرخطی، وابستگی‌های چندبعدی و تعاملات پیچیده میان متغیرها را با دقت بالایی استخراج کنند [۲، ۷، ۱۰، ۴۲]. با این وجود، یکی از چالش‌های اساسی این مدل‌ها، ماهیت «جعبه سیاه» آن‌ها و محدودیت در تفسیرپذیری نتایج است؛ موضوعی که اعتمادپذیری آن‌ها را در تصمیم‌سازی‌های مدیریتی و سیاست‌گذاری کاهش می‌دهد.

برای غلبه بر این محدودیت، در سال‌های اخیر چارچوب‌های ترکیبی توسعه یافته‌اند که با ادغام روش‌های یادگیری ماشینی با تکنیک‌های استخراج و انتخاب ویژگی، ساختارهای پویای نهفته در داده‌های سری زمانی را آشکار می‌سازند. در این میان، بازسازی فضای فاز (Phase Space Reconstruction; PSR) به‌عنوان ابزاری مبتنی بر دینامیک سیستم‌ها، نقش کلیدی در بازنمایی رفتار غیرخطی و غیرایستای سری‌های زمانی ایفا می‌کند. PSR با بازسازی فضای حالت سیستم از طریق متغیرهای تأخیری، امکان استخراج اطلاعات دینامیکی پنهان را فراهم ساخته و پایه‌ای فیزیکی-ریاضی برای بهبود عملکرد مدل‌های پیش‌بینی ایجاد می‌نماید. ترکیب PSR با الگوریتم‌های انتخاب ویژگی نظیر الگوریتم Boruta نیز امکان شناسایی متغیرهای پیش‌بین مؤثر و کاهش افزونگی اطلاعات را فراهم کرده است [۲۰، ۳۱].

در حوزه مدل‌سازی احتمالی، رگرسیون چندکی (Quantile Regression) ابزار قدرتمندی برای تحلیل عدم قطعیت و پیش‌بینی توزیع شرطی متغیر هدف محسوب می‌شود. ادغام رگرسیون چندکی با ساختارهای واین‌کاپولا، چارچوبی انعطاف‌پذیر برای مدل‌سازی همزمان وابستگی چندمتغیره و پیش‌بینی

بازسازی دینامیک سیستم، انتخاب ویژگی، مدل سازی وابستگی چندمتغیره، پیش بینی چندکی و تحلیل عدم قطعیت توسعه داده شد. شکل (۱) نمای کلی مراحل تحقیق را نشان می دهد.

بازسازی فضای فاز (PSR)

بازسازی فضای فاز (PSR) یک تکنیک تحلیل غیرخطی سری های زمانی است که ریشه در نظریه آشوب و سیستم های دینامیکی دارد و امکان بازنمایی دینامیک پنهان یک سیستم را با استفاده از یک سری زمانی تک متغیره فراهم می سازد [۳۳]. براساس قضیه تعبیه تاکنر، رفتار دینامیکی سیستم با ساخت بردارهای مختصات تأخیری مطابق رابطه (۱) بازسازی می شود:

(۱):

$$x'(t) = [x(t), x(t+\tau), x(t+2\tau), \dots, x(t+m-1)\tau]$$

در این رابطه، $x(t)$ سری زمانی مشاهده شده، τ تأخیر زمانی و m بعد تعبیه است. برای بازسازی معتبر دینامیک سیستم، شرط $m > 2D_B$ باید برقرار باشد که در آن D_B بعد شمارش جعبه جاذب است [۲۷، ۳۲]. به منظور تعیین پارامترهای تعبیه، تأخیر زمانی بهینه τ با استفاده از معیار میانگین اطلاعات متقابل (AMI) انتخاب شد؛ به گونه ای که اولین مینیمم منحنی AMI به عنوان نقطه ای استقلال نسبی مؤلفه ها در نظر گرفته شد [۱۱، ۳۹].

انتخاب ویژگی با استفاده از الگوریتم رندوم

فارست

انتخاب ویژگی با استفاده از الگوریتم جنگل تصادفی مبتنی بر ارزیابی میزان اهمیت هر متغیر در بهبود دقت مدل های رگرسیون یا طبقه بندی است. در این رویکرد، مجموعه ای از درخت های تصمیم مستقل داده ها و زیرمجموعه ای از ویژگی ها ساخته می شود. اهمیت هر ویژگی X معمولاً براساس میزان کاهش خطای پیش بینی یا کاهش ناخالصی گینی محاسبه می گردد.

مورد مطالعه حدود ۹۱۳ کیلومتر مربع است [۳۵]. بدین منظور از داده های اقلیمی باران سنجی ایستگاه سینوپتیک بندرعباس در بازه زمانی ۶۲ ساله (۱۹۶۰-۲۰۲۲) استفاده شد و برای تحلیل های آماری و توسعه مدل های پیش بینی مورد استفاده قرار گرفته اند.

به منظور توسعه و ارزیابی مدل های پیش بینی، داده های موجود به صورت زمان محور^۱ به دو مجموعه ی آموزش و آزمایش تقسیم شدند. این تقسیم بندی با هدف شبیه سازی شرایط واقعی پیش بینی و جلوگیری از نشت اطلاعات زمانی^۲ انجام گرفت. بدین صورت که اول دوره ی زمانی (از سال ۱۹۶۰ تا ۲۰۰۳) به عنوان مجموعه ی آموزش و دوره ی پایانی (از سال ۲۰۰۴ تا ۲۰۲۲) به عنوان مجموعه ی آزمایش مستقل در نظر گرفته شد. روش تقسیم زمانی برای داده های سری زمانی اقلیمی از این جهت ضروری است که استقلال واقعی بین نمونه های آموزش و آزمون را حفظ کند و از خوش بین نمایی^۳ معیارهای ارزیابی که ناشی از پیش دستی مدل از اطلاعات آینده است، جلوگیری نماید.

این رویکرد به ویژه در مطالعات پیش بینی های هیدرواقلیمی بلندمدت به عنوان یک استاندارد روش شناختی پذیرفته شده است [۱۶]. در جدول (۱)، مشخصات آماری متغیرهای اصلی شامل دمای میانگین، حداکثر و حداقل دما، تبخیر و تعرق پتانسیل و بارش ارائه شده است. نتایج این جدول نشان دهنده تغییر پذیری قابل توجه پارامترهای اقلیمی در منطقه بندرعباس است که بیانگر شرایط گرم و خشک تا نیمه مرطوب در ماه های مختلف سال می باشد.

روش تحقیق

به منظور مدل سازی دینامیک پیچیده و غیرخطی خشکسالی و کمی سازی عدم قطعیت پیش بینی، در این پژوهش یک چارچوب چندمرحله ای شامل

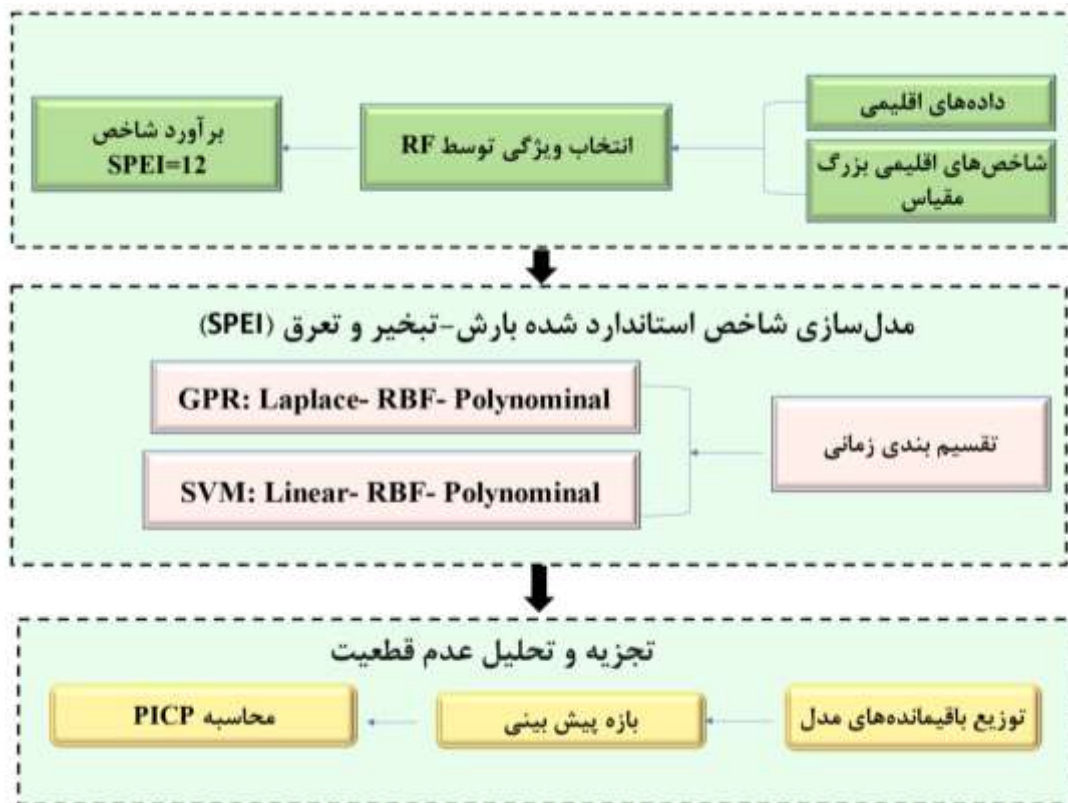
¹ Temporal Split

² Temporal Data Leakage

³ Overoptimism

جدول ۱- مشخصات آماری متغیرهای مورد استفاده

ایستگاه	بارش (میلی‌متر)	تبخیر و تعرق		میانگین رطوبت نسبی	دمای میانگین (C°)	حداکثر دما (C°)
		پتانسیل (میلی‌متر)	حداقل دما (C°)			
حداکثر	۲۶۳	۲۰۰۴	۳۷	۷۴	۴۳/۵	۵۱
حداقل	۰	۱۶۲۶	۱۱	۴۳/۵	۱۸/۵	۲۳
میانگین	۱۴/۱	۱۸۳۹	۲۸/۴	۵۶/۳۹	۳۲/۶	۳۶/۸
انحراف معیار	۳۳/۳	۷۹/۹	۶/۲۵	۵/۱۵	۶/۱	۶/۲۳



شکل ۱- روند نمای تحقیق

$$\text{Importance}(X_j) = \frac{1}{N_T} \sum_{k=1}^{N_T} \sum_{t \in T_k: v(t)=X_j} \Delta I_j(t)$$

که در آن N_T تعداد کل درخت‌ها و $v(t)$ ویژگی مورد استفاده در گره t است. در نهایت، مقادیر اهمیت نرمال‌سازی شده و ویژگی‌هایی با بیشترین اهمیت به‌عنوان ورودی‌های کلیدی مدل انتخاب می‌شوند. این فرآیند ضمن کاهش ابعاد داده، موجب حذف ویژگی‌های کم‌اثر و بهبود کارایی مدل‌های یادگیری ماشینی می‌شود [۱۵].

کاهش ناخالصی گینی حاصل از تقسیم یک گره t با استفاده از ویژگی X_j به‌صورت رابطه (۲) بیان می‌شود:

$$\Delta I_j = p(t)[I(t) - p_L I(t_L) - p_R I(t_R)] \quad (2)$$

که در آن $I(t)$ ناخالصی گره والد، $I(t_L)$ و $I(t_R)$ ناخالصی گره‌های فرزند چپ و راست، و p_L و p_R احتمال تخصیص نمونه‌ها به هر یک از گره‌های فرزند هستند. اهمیت نهایی ویژگی X_j از میانگین مجموع کاهش ناخالصی‌های ناشی از استفاده از آن ویژگی در تمام درخت‌ها محاسبه می‌شود:

$$(3)$$

توابع کاپولا

کاپولا نخستین بار توسط اسکالر در سال ۱۹۵۹ معرفی شد [۲۹]. کاپولا توابعی هستند که توابع توزیع چند متغیره را با توابع توزیع حاشیه‌ای تک‌بعدی متصل می‌کنند. تابع کاپولا می‌تواند ساختار همبستگی بین دو یا چند متغیر تصادفی را به نمایش بگذارد. اخیراً نیز این توابع به‌عنوان روشی عملی و کارآمد برای مدل‌سازی همبستگی داده‌های چند متغیره مطرح شده است [۱۲]. اگر متغیرهای تصادفی $X_1 \dots X_n$ به ترتیب تابع توزیع حاشیه‌ای دلخواه $F_1(x_1) \dots F_n(x_n)$ را دنبال کنند، یک کاپولا، C وجود دارد که این تابع، توزیع حاشیه‌ای را با تابع توزیع $F_1(x_1) \dots F_n(x_n)$ به صورت زیر ترکیب می‌کند:

$$F(x_1) \dots F_n(x_n) = C\{F_1(x_1) \dots F_n(x_n)\} = \quad (۴)$$

$$C(u_1 \dots u_n), x_1 \dots x_n \in \mathbb{R}$$

اگر توزیع‌های حاشیه‌ای $F_i(x_i)$ پیوسته باشند، تابع C منحصر به فرد است. برعکس، اگر C یک تابع K یک بعدی باشد، F یک تابع توزیع N بعدی و $F_1(x_1) \dots F_n(x_n)$ توزیع حاشیه‌ای مربوطه است.

کاپولای واین ابتدا توسط جو^۴ در سال ۱۹۹۶ معرفی شد [۱۷] و سپس توسط آس^۵ و همکاران در سال ۲۰۰۹ توسعه بیشتری یافت [۳]. این کاپولا می‌تواند یک کاپولای متغیر چندبعدی را از طریق (PCCs) به کاپولاهای دوبعدی تجزیه کند، که به‌طور گسترده در تحلیل چند متغیره غیرخطی کاربرد دارند. طبق قضیه اسکالر [۳۷]، تابع توزیع مشترک یک چند متغیره n بعدی، $F(Z_1, \dots, Z_n)$ ، را می‌توان به صورت زیر بیان کرد:

$$F(Z_1, \dots, Z_n) = C\{F_1(z_1), \dots, F_n(z_n)\} = \quad (۵)$$

$$(u_1, \dots, u_n)$$

که در آن:

F توابع توزیع حاشیه‌ای برای هر متغیر هستند.

C یک تابع کاپولا است که وابستگی بین متغیرها را مستقل از حاشیه‌ها توصیف می‌کند.

در این پژوهش از کاپولای واین و ساختار کوانتای آن (QVineCopula) برای بازنمایی وابستگی‌های

چندمتغیره و ترکیب آن با چارچوب بازسازی فضای فاز (PSR) در مدل‌سازی دقیق و تفسیرپذیر خشکسالی استفاده شد.

رگرسیون چندکی

رگرسیون چندکی [۲۱] برای برآورد چندک‌های شرطی متغیر پاسخ به‌کار می‌رود و برخلاف روش حداقل مربعات، کل توزیع شرطی داده‌ها و رفتارهای حدی را مدل‌سازی می‌کند. تابع زیان چندکی (پین‌بال) به صورت زیر تعریف می‌شود:

$$\rho_q(u) = \begin{cases} qu & u \geq 0 \\ (q-1)u & u < 0 \end{cases} \quad (۶)$$

که در آن $u = y - \hat{y}$ و q سطح چندک است. امید ریاضی این تابع به شکل زیر بیان می‌شود:

$$(۷)$$

$$E[\rho_q(y - \hat{y})] = \int_{-\infty}^{\hat{y}} (1-q)(\hat{y} - y)f(y) dy + \int_{\hat{y}}^{\infty} q(y - \hat{y})f(y) dy$$

با مشتق‌گیری نسبت به \hat{y} و برابر قرار دادن آن با صفر، شرط بهینگی حاصل می‌شود:

$$F(\hat{y}) = q \quad (۸)$$

که نشان می‌دهد مقدار بهینه \hat{y} ، برابر با چندک q ام توزیع متغیر پاسخ است. بنابراین، کمینه‌سازی تابع زیان پین‌بال منجر به برآورد مستقیم چندک‌های شرطی و تحلیل مؤثر عدم قطعیت می‌شود.

الگوریتم تقویت گرادیان شدید و نسخه چندکی

آن

الگوریتم تقویت گرادیان شدید (Extreme Gradient Boosting - XGB) یکی از دقیق‌ترین روش‌های یادگیری ماشین برای تحلیل داده‌های پیچیده و بزرگ‌مقیاس است که بر پایه‌ی ترکیب افزایشی درخت‌های تصمیم و بهینه‌سازی گرادیانی توسعه یافته است [۵،۴۶]. در این روش، مدل به صورت ترتیبی ساخته شده و در هر مرحله، یادگیرنده‌ی جدید خطای مدل قبلی را کاهش

⁴ Joe

⁵ Aas

جنگل تصادفی و جنگل رگرسیون چندکی (RF و QRF)

الگوریتم جنگل تصادفی (Random Forest – RF) یک روش یادگیری نظارت‌شده مبتنی بر تجمیع درخت‌های تصمیم تصادفی است که توانایی بالایی در مدل‌سازی روابط غیرخطی و پیچیده دارد [۴، ۱۴]. تصادفی‌سازی داده‌ها و ویژگی‌ها موجب کاهش بیش‌برازش و افزایش پایداری پیش‌بینی‌ها می‌شود [۲۷].

در فرآیند رشد درخت‌ها، معیار تقسیم گره با استفاده از شاخص ناخالصی تعریف می‌شود:

$$L_G(t_x(x_i)) = 1 - \sum_{j=1}^m f(t_x(x_i), j)^2 \quad (۱۳)$$

که در آن $f(t_x(x_i), j)$ نسبت نمونه‌های اختصاص‌یافته به شاخه j است. در مسائل رگرسیونی، خروجی نهایی مدل از میانگین پیش‌بینی تمام درخت‌ها به دست می‌آید:

$$\hat{y} = \frac{1}{K} \sum_{k=1}^K h(x, \theta_k) \quad (۱۴)$$

با وجود دقت بالای RF در تخمین مقدار میانگین، این مدل اطلاعاتی از عدم قطعیت پیش‌بینی ارائه نمی‌دهد. به منظور رفع این محدودیت، جنگل رگرسیون چندکی (Quantile Regression Forest – QRF) توسعه یافته است [۲۳، ۶۰]. در QRF، به جای میانگین، توزیع شرطی کامل متغیر پاسخ در برگ‌های درخت ذخیره می‌شود و تابع توزیع شرطی به صورت زیر برآورد می‌گردد:

$$\hat{F}(y | x) = \sum_{i=1}^n w_i(x) I(y_i \leq y) \quad (۱۵)$$

چندک مورد نظر از معکوس این تابع به دست می‌آید:

$$\hat{Q}_\tau(x) = \hat{F}^{-1}(\tau | x) \quad (۱۶)$$

به این ترتیب، QRF ضمن حفظ دقت RF، امکان برآورد چندک‌ها و تحلیل عدم قطعیت پیش‌بینی را فراهم می‌سازد و به‌ویژه برای مدل‌سازی پدیده‌های اقلیمی و هیدرولوژیکی مناسب است.

روش تحلیل عدم قطعیت UNEEC

روش UNEEC (تخمین عدم قطعیت با استفاده از خوشه‌بندی خطا) یک رویکرد ناپارامتریک برای کمی‌سازی عدم قطعیت در مدل‌های پیش‌بین است

می‌دهد. تابع پیش‌بینی در مرحله t به صورت زیر بیان می‌شود:

$$(۹)$$

$$f_i^{(t)} = \sum_{k=1}^t f_k(x_i) = f_i^{(t-1)} + f_t(x_i)$$

تابع هدف XGB شامل تابع زیان و عبارت منظم‌سازی است:

$$(۱۰)$$

$$obj^{(t)} = \sum_{i=1}^n l(\hat{y}_i, y_i) + \sum_{k=1}^t \Omega(f_k)$$

که در آن، l تابع زیان و Ω عبارت منظم‌سازی بوده و به صورت زیر تعریف می‌شود:

$$\Omega(f) = \frac{1}{2} \lambda \|W\|^2 + \gamma T \quad (۱۱)$$

در این رابطه، W بردار امتیازهای برگ، λ پارامتر منظم‌سازی و γ حداقل زیان لازم برای تقسیم گره است. این ساختار با کنترل پیچیدگی مدل، از بیش‌برازش جلوگیری کرده و پایداری پیش‌بینی را افزایش می‌دهد.

نسخه‌ی چندکی این الگوریتم، یعنی تقویت گرادیان شدید چندکی (QXGB)، برای تخمین چندک‌های شرطی و ساخت فواصل پیش‌بینی توسعه یافته و ترکیبی از رگرسیون چندکی و XGB است. در این روش، به جای تابع زیان میانگین مربعات خطا، از تابع زیان قدر مطلق وزن‌دار استفاده می‌شود. با توجه به غیرمشتق‌پذیر بودن این تابع در مبدأ، برای پیاده‌سازی در چارچوب XGB، از تابع هابر به عنوان تقریب نرم استفاده می‌شود [۶، ۱۸]:

$$(۱۲)$$

$$L_\delta(x) = \begin{cases} \frac{1}{2} x^2 & |x| \leq \delta \\ \delta \left(|x| - \frac{1}{2} \delta \right) & |x| > \delta \end{cases}$$

این تقریب امکان استفاده از بهینه‌سازی مبتنی بر گرادیان را فراهم کرده و QXGB را قادر می‌سازد ضمن حفظ توان پیش‌بینی XGB، ناپاطمینانی و دامنه‌ی تغییرپذیری خروجی را نیز مدل‌سازی کند [۴۱].

نتایج

شواهد وجود ساختار دینامیکی و ضرورت

بازسازی فضای فاز (PSR)

شکل ۲ (الف) نمودار میانگین اطلاعات متقابل (Average Mutual Information; AMI) را برای سری زمانی SPEI12 نشان می‌دهد که میزان وابستگی اطلاعاتی بین مقادیر سری زمانی در فواصل زمانی مختلف را ارزیابی می‌کند. مقادیر بالای AMI در تأخیرهای کوتاه بیانگر اشتراک اطلاعاتی زیاد بین نمونه‌های زمانی مجاور است. با افزایش تأخیر زمانی (τ) ، مقدار AMI کاهش یافته و نشان‌دهنده افت تدریجی وابستگی اطلاعاتی بین داده‌هاست. نخستین مینیمم محلی مشخص در بازه $5-10 \approx \tau$ مشاهده می‌شود که نشان‌دهنده کاهش اطلاعات تکراری در عین حفظ ساختار پویای سیستم است. بنابراین، این بازه به‌عنوان تأخیر زمانی بهینه برای بازسازی فضای فاز انتخاب شد.

شکل ۲ (ب) نتایج تحلیل بُعد تعبیه (Embedding Dimension) مبتنی بر روش همسایگان کاذب (False Nearest Neighbors) را نمایش می‌دهد. شاخص $E1(d)$ با افزایش بُعد تعبیه، روندی کاهشی داشته و در بعد $d=2$ به حد آستانه تعیین‌شده می‌رسد. هم‌زمان، مقدار $E2(d)$ نیز در همین بعد به صفر میل می‌کند. این همگرایی نشان می‌دهد که افزایش بُعد تعبیه فراتر از $d=2$ به بهبود معناداری در جداسازی نقاط منجر نمی‌شود.

این نتایج به‌طور صریح تأیید می‌کنند که سری زمانی SPEI12 یک فرآیند تصادفی بدون حافظه نیست، بلکه دارای رفتار دینامیکی و حافظه زمانی است. بنابراین، استفاده از PSR پیش‌نیاز ضروری برای استخراج ویژگی‌های سازگار با ساختار فیزیکی سامانه خشکسالی و توسعه مدل‌های یادگیری ماشین قابل‌اتکا محسوب می‌شود.

که بدون فرض خاصی درباره توزیع خطا، از داده‌های تاریخی بهره می‌برد [۲۸]. در این روش، پس از آموزش مدل بر روی داده‌های تاریخی و محاسبه خطاهای پیش‌بینی براساس رابطه زیر:

$$e_i = Y_i - \hat{Y}_i \quad (17)$$

که در آن Y_i مقدار واقعی و \hat{Y}_i مقدار پیش‌بینی شده است، فضای ورودی X به خوشه K خوشه (C_1, C_2, \dots, C_K) تقسیم می‌شود (به‌عنوان مثال با روش (k-means) برای هر خوشه C_j ، میانگین و انحراف معیار خطاها به‌ترتیب از روابط زیر به‌دست می‌آیند:

$$\bar{e}_j = \frac{1}{N_j} \sum_{i \in C_j} e_i \quad (17)$$

$$\sigma_j = \sqrt{\frac{1}{N_j - 1} \sum_{i \in C_j} (e_i - \bar{e}_j)^2}$$

که در آن N_j تعداد نمونه‌ها در خوشه j است.

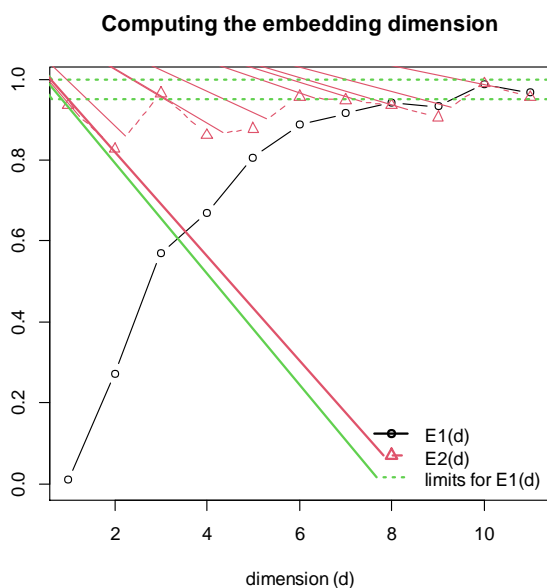
سپس برای هر خوشه، بازه اطمینان پیش‌بینی با استفاده از کوانتیل‌های پایین و بالای توزیع خطا ($q_{1-\alpha/2}$ و $q_{\alpha/2}$) به‌شکل زیر ساخته می‌شود:

$$\hat{Y}_i^L, \hat{Y}_i^U = \hat{Y}_i + q_{\alpha/2}, q_{1-\alpha/2} \quad (19)$$

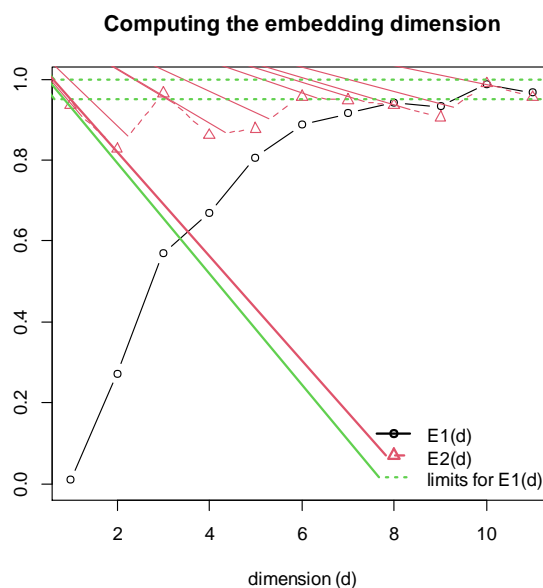
دقت بازه‌های پیش‌بینی با شاخص PICP (Prediction Interval Coverage Probability) ارزیابی می‌شود:

$$PICP = \frac{1}{N} \sum_{i=1}^N I(Y_i \in [\hat{Y}_i^L, \hat{Y}_i^U]) \quad (20)$$

که در آن $I(\cdot)$ تابع نشانگر است (۱ اگر Y_i در بازه پیش‌بینی قرار گیرد، در غیر این صورت صفر این رویکرد، با در نظر گرفتن توزیع خطا در خوشه‌های مختلف، عدم قطعیت پیش‌بینی را به‌صورت داده‌محور و بدون فرض پارامتریک برآورد می‌کند [۴۳] شکل ۱ روندنمای تحقیق را نشان می‌دهد.



ب



الف

شکل ۲- الف: مقدار بهینه τ برای SPEI12 و ب: نمودار بعد تعبیه

در مورد متغیرهای حرارتی، از جمله دمای بیشینه، کمینه و میانگین (t_{max_max} ، t_{max_min} و t_{mean})، رابطه‌ای غالباً منفی و غیرخطی با SPEI12 مشاهده می‌شود. افزایش دما، به‌ویژه در شرایط خشکسالی شدید، منجر به افزایش تبخیر-تعرق بالقوه و در نتیجه کاهش مقادیر SPEI12 می‌شود. فاصله قابل توجه بین منحنی‌های کوانتایلی در این متغیرها نشان می‌دهد که اثر دما بر خشکسالی در شرایط حدی به مراتب قوی‌تر از شرایط نرمال است.

متغیرهای مرتبط با رطوبت نسبی (u_{max} و u_{mean}) رابطه‌ای مثبت با SPEI12 نشان می‌دهند، به طوری که افزایش رطوبت نسبی به بهبود شرایط رطوبتی و کاهش شدت خشکسالی منجر می‌شود. این اثر در کوانتایل‌های بالاتر ($\alpha = 0.9$) بارزتر است که بیانگر نقش تعدیل‌کننده رطوبت در شرایط ترسالی می‌باشد. در مقابل، سرعت باد (w_{in}) وابستگی ضعیف‌تر اما غیرخطی با SPEI12 نشان می‌دهد و عمدتاً به‌عنوان یک عامل تشدیدکننده ثانویه از طریق افزایش تبخیر-تعرق عمل می‌کند.

به‌طور کلی، نتایج تحلیل واین کاپولا نشان می‌دهد که خشکسالی بلندمدت در منطقه مورد مطالعه

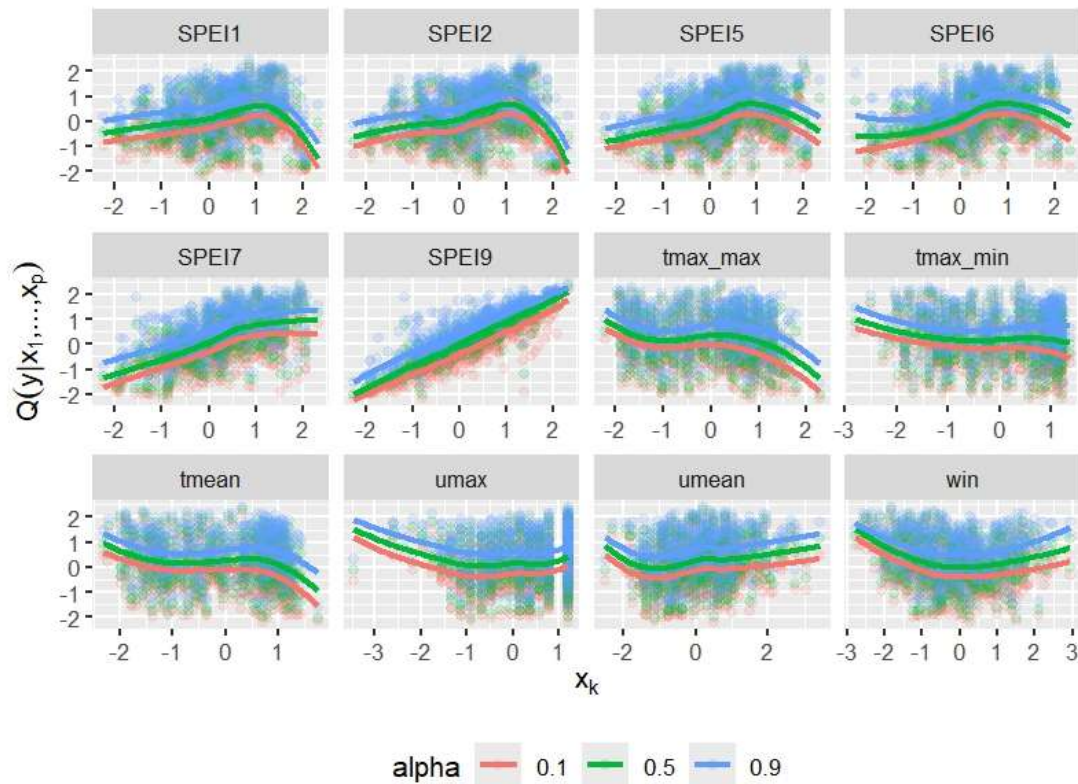
ساختار وابستگی‌های فیزیکی-آماري متغیرهای اقلیمی (تحلیل Vine Copula)

شکل ۳ ساختار وابستگی چندمتغیره بین SPEI12 و مجموعه‌ای از متغیرهای اقلیمی را براساس تحلیل واین کاپولا در سطوح مختلف کوانتایلی (۰.۵ و ۰.۹ α و $0.1 = \alpha$) نشان می‌دهد. نتایج بیانگر وجود روابط غیرخطی، نامتقارن و وابسته به شرایط حدی بین SPEI12 و متغیرهای توضیحی است، به طوری که شدت و جهت وابستگی در شرایط خشکسالی شدید، نرمال و ترسالی تفاوت معناداری دارد.

شاخص‌های خشکسالی با مقیاس‌های زمانی متفاوت (SPEI1، SPEI16، SPEI17 و SPEI19) بیشترین و پایدارترین وابستگی را با SPEI12 نشان می‌دهند. این وابستگی مثبت و قوی بیانگر وجود حافظه زمانی در سیستم هیدرولوژیکی منطقه است، به گونه‌ای که شرایط رطوبتی و خشکسالی در مقیاس‌های زمانی مجاور نقش تعیین‌کننده‌ای در تداوم و تشدید خشکسالی بلندمدت ایفا می‌کنند. شدت این وابستگی در کوانتایل‌های پایین ($\alpha = 0.1$) بیشتر است که نشان‌دهنده حساسیت بالاتر SPEI12 به شرایط حدی خشک می‌باشد.

وابستگی پیچیده، مبنای مناسبی برای انتخاب متغیرهای ورودی در مدل‌های پیش‌بینی مبتنی بر یادگیری ماشین فراهم می‌کند و بر ضرورت استفاده از روش‌های وابستگی محور مانند Vine Copula در تحلیل خشکسالی تأکید دارد.

حاصل برهم‌کنش همزمان و غیرخطی بین شاخص‌های خشکسالی کوتاه‌مدت (به‌عنوان عامل کنترل‌کننده تداوم)، متغیرهای حرارتی (به‌عنوان عامل تشدیدکننده) و متغیرهای رطوبتی و دینامیکی جو (به‌عنوان عوامل تعدیل‌کننده) است. این ساختار



شکل ۳- ساختار وابستگی چندمتغیره بین SPEI12 و متغیرهای اقلیمی حاصل از تحلیل واین کاپولا

الگو، به‌وضوح نشان‌دهنده نقش محوری و سلسله‌مراتبی شاخص‌های خشکسالی با تاخیر زمانی کوتاه‌تر در پیش‌بینی شرایط خشکسالی بلندمدت‌تر است. به بیان دیگر، برای پیش‌بینی دقیق‌تر خشکسالی ۱۲ ماهه، مدل بیشترین وابستگی را به اطلاعات خشکسالی در ماه نهم قبل داشته است.

متغیرهای SPEI2 تا SPEI7 و نیز SPEI1 اهمیت متوسط و نسبتاً مشابهی از خود نشان دادند که حاکی از مشارکت جمعی و تدریجی شرایط رطوبتی در مقیاس‌های زمانی مختلف در شکل‌گیری خشکسالی بلندمدت است. در مقابل، دو متغیر اقلیمی میانگین دمای ماهانه (Tmean) و متغیر WS (سرعت باد) با کمترین میزان اهمیت (به‌ترتیب

انتخاب متغیرهای غالب و اثر حافظه زمانی

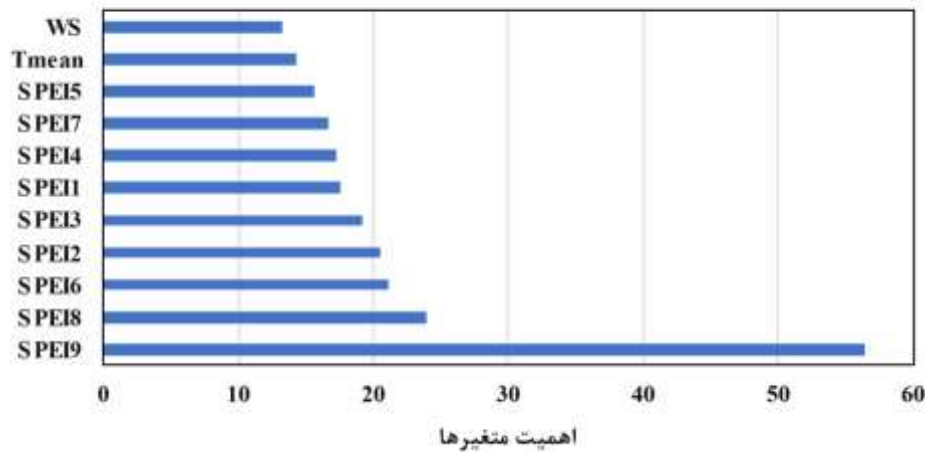
سامانه (RF Feature Selection)

نتایج ارائه‌شده در شکل ۴ بیانگر اهمیت نسبی متغیرها در پیش‌بینی SPEI12 براساس الگوریتم جنگل تصادفی است. هم‌پوشانی بازه $5-10 \sim \tau$ با اهمیت SPEI9 نشان می‌دهد که تأخیر بهینه استخراج‌شده از AMI با حافظه مؤثر سیستم در مدل‌های یادگیری ماشین سازگار است. این یافته حاکی از آن است که وضعیت خشکسالی در بازه زمانی ۹ ماهه، قوی‌ترین پیش‌بینی‌کننده برای وضعیت خشکسالی ۱۲ ماهه (متغیر وابسته) در منطقه مورد مطالعه می‌باشد. پس از آن، متغیرهای SPEI8 و SPEI6 در رتبه‌های بعدی قرار دارند. این

تاریخچه تجمعی وضعیت خشکسالی) که در مقادیر SPEI با تاخیرهای مختلف انعکاس یافته (تکیه کرده است تا بر وضعیت لحظه‌ای پارامترهای جوی. در مجموع، خروجی شکل ۴ این فرضیه را تقویت می‌کند که پویایی خشکسالی هیدروکلیماتولوژیک در منطقه مورد مطالعه از حافظه بلندمدت قوی برخوردار است و خشکسالی در مقیاس ۹ ماهه می‌تواند به‌عنوان یک نشانگر پیش‌آگهی مؤثر برای پایش و هشدار زودهنگام خشکسالی در مقیاس سالانه (۱۲ ماهه) مورد توجه قرار گیرد.

۲۸/۱۴ و ۲۴/۱۳) شناسایی شدند. این نتیجه می‌تواند نشان‌دهنده دو مکانیسم باشد: نخست، اثرات مستقیم پارامترهایی مانند دما و بارش یا باد، عمدتاً از طریق شاخص ترکیبی SPEI که خود تابعی از بارش و تبخیر و تعرق بالقوه (متأثر از دما) است، در مدل تبیین شده‌اند. دوم، این یافته بر این اصل هیدرواقليمی تاکید دارد که خشکسالی یک پدیده خزنده و تجمعی است؛ بنابراین شرایط نرمال یا حتی مرطوب کوتاه‌مدت نمی‌تواند به‌سرعت اثرات انباشته شده یک دوره خشکسالی بلندمدت را خنثی کند. از این رو، مدل رندوم فارست برای پیش‌بینی SPEI12، بیشتر بر

Feature selection



شکل ۴- نمودار اهمیت متغیرها با استفاده از روش RF

Copula نیز با مقدار R^2 برابر ۰/۰۹ ضعیف‌ترین عملکرد را در مرحله آموزش نشان می‌دهد. با انتقال به داده‌های آزمون، افت عملکرد در همه مدل‌ها مشهود است. ضرایب تبیین به ۰/۱۱ برای RF، ۰/۲۰ برای XGB و مقدار بسیار پایین ۰/۰۱ برای Vine Copula کاهش یافته‌اند. همزمان، مقادیر RMSE و MAE افزایش چشمگیری داشته‌اند (به‌ویژه در مدل Vine Copula با MAE برابر ۱/۰۰)، که بیانگر ناتوانی مدل‌ها در تعمیم‌پذیری و بروز بیش‌برازش شدید در نبود بازنمایی مناسب از ساختار زمانی داده‌هاست. این نتایج نشان می‌دهد که استفاده مستقیم از داده‌های خام سری‌زمانی، بدون

ارزیابی مدل‌های یادگیری ماشین بدون و با بازسازی فضای فاز

نتایج ارزیابی عملکرد مدل‌های یادگیری ماشین در دو حالت «بدون بازسازی فضای فاز (PSR)» و «با بازسازی فضای فاز» در جدول ۲ ارائه شده است. در حالت بدون PSR، عملکرد تمامی مدل‌ها در مرحله آموزش محدود و در مرحله آزمون ضعیف است. مدل Random Forest (RF) و XGBoost در مرحله آموزش به‌ترتیب ضرایب تبیین ۰/۱۹ و ۰/۲۳ را ثبت کرده‌اند که نشان‌دهنده توانایی محدود آن‌ها در توضیح تغییرپذیری داده‌های هدف است. مدل Vine

می‌دهد. این مدل در مرحله آموزش به R^2 برابر $0/83$ دست یافته و در مرحله آزمون نیز مقدار بالای $R^2 = 0/791$ را همراه با کمترین مقادیر خطا ($RMSE = 0/38$ و $MAE = 0/26$) حفظ کرده است. فاصله نسبتاً اندک بین نتایج آموزش و آزمون نشان‌دهنده پایداری بالا و عدم وابستگی مدل به برازش صرف داده‌های آموزشی است. برتری PSR-Vine-Copula را می‌توان به توانایی آن در مدل‌سازی وابستگی‌های غیرخطی پیچیده در کنار بازنمایی دینامیکی داده‌ها توسط PSR نسبت داد.

در مجموع، نتایج جدول ۲ به‌وضوح نشان می‌دهد که بازسازی فضای فاز (PSR) یک گام پیش‌پردازشی ضروری برای ارتقای دقت و قابلیت تعمیم مدل‌های یادگیری ماشین در پیش‌بینی خشکسالی است. در حالی که مدل‌های بدون PSR دچار بیش‌برازش شده و برای کاربرد عملی نامعتبر هستند، ادغام PSR با مدل‌های یادگیری ماشین، به‌ویژه Vine Copula، چارچوبی پایدارتر و واقع‌بینانه‌تر برای پیش‌بینی خشکسالی فراهم می‌کند.

در نظر گرفتن پویایی و حافظه زمانی سامانه اقلیمی، برای پیش‌بینی قابل اتکای خشکسالی کافی نیست. پس از اعمال بازسازی فضای فاز (PSR) و انتخاب ویژگی با استفاده از Random Forest، بهبود قابل‌توجهی در عملکرد تمامی مدل‌ها مشاهده می‌شود. در مرحله آموزش، مقدار R^2 مدل‌های RF-PSR-RF و RF-PSR-XGB به ترتیب به $0/59$ و $0/51$ افزایش یافته و مقادیر RMSE و MAE به‌طور محسوسی کاهش پیدا کرده‌اند. این نتایج بیانگر آن است که PSR با آشکارسازی ساختار دینامیکی نهفته در سری زمانی، یادگیری مؤثرتری از روابط سیستم را ممکن ساخته است.

مهم‌تر آنکه این بهبود به داده‌های آموزش محدود نمانده و در مرحله آزمون نیز حفظ شده است. در این مرحله، ضرایب تبیین RF-PSR-RF و RF-PSR-XGB به ترتیب برابر $0/487$ و $0/49$ ثبت شده که نشان‌دهنده افزایش قابل‌توجه پایداری و قابلیت تعمیم مدل‌ها نسبت به حالت بدون PSR است.

در میان تمامی مدل‌ها، چارچوب ترکیبی PSR-Vine-Copula عملکردی متمایز و برتر ارائه

جدول ۲- ارزیابی مدل در حالت با و بدون بازسازی فضای فاز

Training			Testing			
Without PSR						
Model	R^2	RMSE	MAE	R^2	RMSE	MAE
RF	0.19	0.41	0.33	0.11	1.164	0.981
XGB	0.23	0.37	0.24	0.20	1.66	0.942
VineCopula	0.09	0.86	0.70	0.01	1.165	1.00
RF-PSR						
RF-PSR-RF	0.59	0.16	0.12	0.487	0.65	0.53
RF-PSR-XGB	0.51	0.15	0.13	0.49	0.66	0.51
PSR-Vine-Copula	0.83	0.12	0.10	0.791	0.38	0.26

«PSR» و «با بازسازی فضای فاز» به‌روشنی نشان می‌دهد که پیش‌پردازش مناسب داده‌های سری زمانی نقش تعیین‌کننده‌ای در بهبود دقت و قابلیت تعمیم‌پذیری مدل‌ها دارد. در حالت بدون PSR،

ارزیابی مدل‌های یادگیری ماشین چندکی بدون و با بازسازی فضای فاز نتایج ارزیابی عملکرد مدل‌های یادگیری ماشین چندکی در دو سناریوی «بدون بازسازی فضای فاز

یادگیری روابط پایدار را برای مدل‌های چندکی ممکن ساخته است.

در این میان، مدل ترکیبی PSR-QVineCopula عملکردی برتر و متمایز ارائه کرد. این مدل در مرحله آموزش به R^2 برابر 0.83 دست یافت و در مرحله آزمون نیز مقدار بالای $R^2 = 0.80$ را همراه با کمترین خطا ($RMSE = 0.38$ و $MAE = 0.26$) حفظ کرد. فاصله اندک بین نتایج آموزش و آزمون در این مدل بیانگر پایداری و قابلیت اعتماد بالای آن است. برتری PSR-QVineCopula را می‌توان به توانایی مدل‌سازی وابستگی‌های غیرخطی و پیچیده در چارچوب کاپولا، در کنار بازنمایی ساختار دینامیکی داده‌ها توسط PSR، نسبت داد.

به‌طور کلی، مقایسه دو سناریو نشان می‌دهد که استفاده از PSR به‌عنوان مرحله پیش‌پردازش، داده‌های خام سری‌زمانی را به یک نمایش ساختاریافته و معنادار تبدیل می‌کند که موجب ارتقای قابل‌توجه دقت، پایداری و قابلیت تعمیم مدل‌ها می‌شود. همچنین، ترکیب PSR با واین‌کاپولا چندکی، با بهره‌گیری هم‌زمان از نظریه سیستم‌های دینامیکی و تحلیل وابستگی چندمتغیره، چارچوبی منسجم و قدرتمند برای پیش‌بینی‌های احتمالی خشکسالی در سامانه‌های اقلیمی پیچیده ارائه می‌دهد.

عملکرد همه مدل‌ها شامل QRF، QXGB و QVineCopula در هر دو مرحله آموزش و آزمون ضعیف بوده است. ضرایب تبیین (R^2) در داده‌های آموزش بین 0.09 تا 0.21 مشاهده شد که نشان‌دهنده توان محدود این مدل‌ها در توضیح ساختار تغییرات داده‌هاست. با انتقال به داده‌های آزمون، این مقادیر به‌طور محسوسی کاهش یافت و در مورد QVineCopula به 0.19 رسید. همچنین، شاخص‌های خطا ($RMSE$ و MAE) در مرحله آزمون افزایش قابل توجهی پیدا کردند، که بیانگر بروز بیش‌برازش و ناتوانی مدل‌ها در یادگیری روابط پایدار و تعمیم آن‌ها به داده‌های مستقل است.

به‌کارگیری روش بازسازی فضای فاز (PSR) همراه با انتخاب ویژگی مبتنی بر جنگل تصادفی باعث بهبود چشمگیر عملکرد مدل‌ها شد. در مرحله آموزش، ضرایب تبیین مدل‌های RF-PSR-QRF و RF-PSR-QXGB به ترتیب به 0.72 و 0.79 رسید و مقادیر خطا کاهش قابل توجهی یافت. نکته مهم‌تر آن است که این بهبود در داده‌های آزمون نیز حفظ شد، به‌گونه‌ای که R^2 به ترتیب برابر با 0.55 و 0.57 ثبت شد و شاخص‌های $RMSE$ و MAE در مقایسه با حالت بدون PSR تقریباً نصف شدند. این نتایج نشان می‌دهد که PSR با بازنمایی حالت‌های دینامیکی نهفته و ساختار زمانی سیستم اقلیمی،

جدول ۳- ارزیابی مدل‌های چندکی در حالت با و بدون بازسازی فضای فاز

Model	Training			Testing		
	R^2	RMSE	MAE	R^2	RMSE	MAE
Without PSR						
QRF	0.20	0.33	0.10	0.13	1.17	0.97
QXGB	0.21	0.43	0.30	0.21	1.26	1.05
QVineCopula	0.09	0.86	0.70	0.19	1.27	1.01
RF-PSR						
RF-PSR-QRF	0.72	0.16	0.12	0.55	0.61	0.50
RF-PSR-QXGB	0.79	0.16	0.11	0.57	0.60	0.49
PSR-QVineCopula	0.83	0.15	0.10	0.80	0.38	0.26

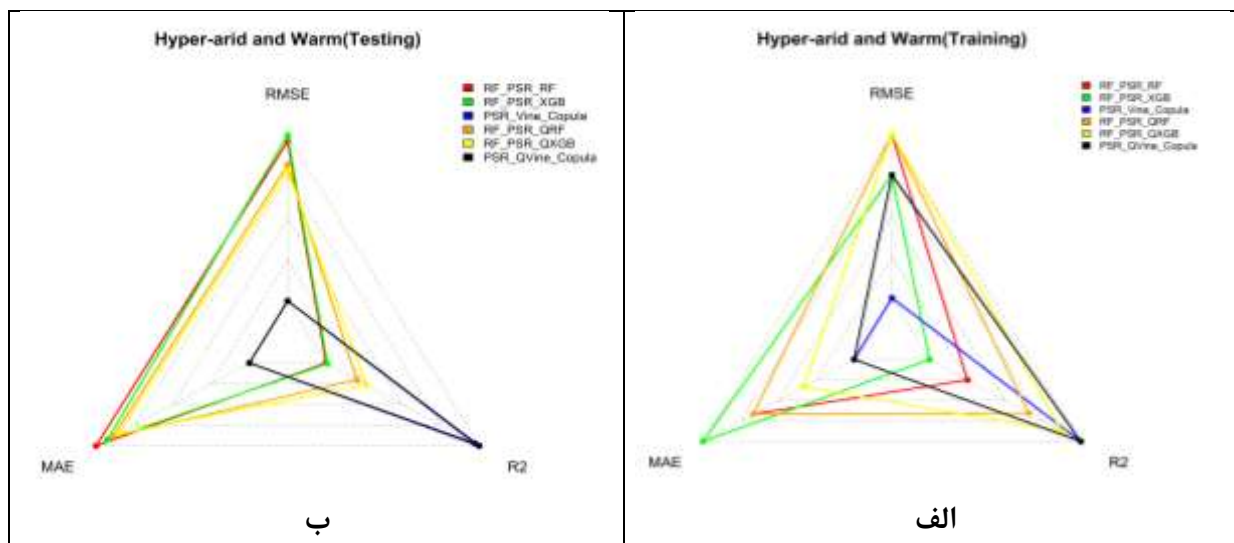
مقایسه مدل‌های یادگیری ماشین و چندکی

در این پژوهش، برای مقایسه‌ی کیفی و کمی عملکرد مدل‌ها از نمودارهای راداری (Radar Plot) استفاده شد. در این نمودارها، هر محور نشان‌دهنده‌ی یکی از شاخص‌های ارزیابی عملکرد مدل شامل RMSE، MAE و R^2 است. از آن‌جا که برای شاخص‌های خطا (MAE و RMSE) مقادیر کمتر نشان‌دهنده‌ی عملکرد بهتر است، جهت این محورها معکوس شد تا در تمامی محورها افزایش مقدار نشان‌دهنده‌ی بهبود عملکرد مدل باشد. در نتیجه، مساحت چندضلعی (یا مثلث راداری) تشکیل‌شده توسط هر مدل بیانگر توان کلی آن در بازتولید و پیش‌بینی دقیق متغیر هدف است؛ به عبارت دیگر، هرچه سطح پوشش مدل در نمودار راداری بزرگ‌تر باشد، آن مدل از نظر شاخص‌های آماری عملکرد دقیق‌تر و پایدارتری دارد.

در مرحله‌ی آموزش (شکل ۵- الف)، مدل‌های کوانتالیی نسبت به مدل‌های کلاسیک عملکرد بهتری نشان دادند؛ زیرا انطباق دقیق‌تری میان مقادیر پیش‌بینی‌شده و داده‌های واقعی ایجاد کردند. در میان همه مدل‌ها، نسخه‌ی PSR-QVineCopula از نظر شاخص‌های دقت (R^2 بالا) و خطا (RMSE و MAE پایین) بهترین نتیجه را داشت. ترکیب روش کاپولا با رویکرد کوانتالیی سبب پوشش بهتر نوسانات داده شد، و به کارگیری مؤلفه‌ی PSR موجب افزایش

پایداری و هم‌خوانی داخلی مدل گردید. به طور کلی، در مرحله‌ی آموزش می‌توان نتیجه گرفت که ساختارهای غیرخطی مدل Vine Copula همراه با کمی‌سازی عدم قطعیت، توانایی بسزایی در بازسازی رفتار اقلیمی داشتند. ترکیب کاپولا با چارچوب کوانتالیی موجب بازنمایی بهتر نوسانات و توزیع SPEI شد، در حالی که PSR پایداری ساختاری مدل را تضمین کرد.

در مرحله‌ی آزمون (شکل ۵- ب) نیز همین الگو حفظ شد و مدل PSR-QVineCopula همچنان بهترین عملکرد را در تعمیم‌پذیری به داده‌های جدید نشان داد. این مدل کمترین افت در عملکرد بین دو مرحله‌ی Training و Testing داشت که بیانگر قدرت بالای مدل در اجتناب از بیش‌برازش است. در مقایسه با نسخه‌های بدون PSR، مدل‌های دارای PSR مقادیر خطا (RMSE و MAE) کمتر و ضریب تعیین (R^2) بالاتری ارائه کردند؛ در نتیجه تأثیر PSR در بهبود پایداری و کاهش خطای پیش‌بینی کاملاً مشهود است. در مجموع، نتایج نمودارهای راداری به طور بصری و کمی تأیید می‌کند که ترکیب PSR، Vine Copula و رگرسیون چندکی دقیق‌ترین، پایدارترین و قابل‌اعتمادترین چارچوب برای پیش‌بینی شاخص SPEI محسوب می‌شود.



شکل ۵- رادار پلات مدل‌های مورد استفاده در فاز آموزش (الف)، فاز آزمایش (ب)

تحلیل عدم قطعیت مدل‌ها

نتایج ارائه‌شده در جدول ۴ و شکل ۶ باندهای بالا و پایین پیش‌بینی (UB و LB)، عرض عدم قطعیت (UA) و شاخص پوشش فاصله پیش‌بینی (PICP) را برای مدل‌های مختلف در دو حالت بدون و با بازسازی فضای فاز (PSR) در مراحل آموزش و آزمون نشان می‌دهد. تحلیل نتایج در دو بلوک مجزا—مدل‌های استاندارد (غیر چندکی) و مدل‌های چندکی—انجام شده است.

در حالت پایه و بدون استفاده از PSR، مدل‌های RF، XGB و Vine Copula عملکرد ضعیفی در برآورد عدم قطعیت دارند. در مرحله آموزش، مقادیر UB و LB فاصله قابل توجهی از صفر دارند (برای مثال، RF با $UB=1.67$ و $LB=-0.14$ و Vine Copula با $UB=1.56$ و $LB=-1.26$) که نشان‌دهنده وجود آریب سیستماتیک است. همزمان، عرض عدم قطعیت نسبتاً بزرگ بوده (UA بین ۱/۱۰ تا ۲/۸۳) و شاخص پوشش در سطح پایینی قرار دارد (PICP بین ۰/۱۵ تا ۰/۴۵).

این ضعف در مرحله آزمون تشدید می‌شود؛ به طوری که کران پایین پیش‌بینی در تمامی مدل‌ها به حدود ۱/۹- نزدیک شده است (مانند $LB=-1.91$ در RF و $LB=-1.90$ در Vine Copula) و دامنه عدم قطعیت به بیش از ۲/۶ واحد افزایش یافته است ($UA=2.68$ برای RF، $UA=2.59$ برای XGB و $UA=2.96$ برای Vine Copula). با وجود این بازه‌های گسترده، مقادیر PICP همچنان پایین باقی‌مانده و در برخی موارد به حداقل ۰/۱۲ کاهش یافته است که بیانگر فاصله‌های پیش‌بینی پهن اما غیرقابل اعتماد است.

با اعمال PSR، بهبود محسوسی در مدل‌های استاندارد RF و XGB مشاهده می‌شود. مقادیر UB و LB به صفر نزدیک‌تر شده و عرض عدم قطعیت کاهش یافته است (برای مثال، UA مدل RF-PSR-

RF در آزمون به ۱/۷۸ رسیده است). با این حال، شاخص پوشش (PICP) در این مدل‌ها همچنان در محدوده متوسط (۰/۵۴ تا ۰/۶۴) باقی می‌ماند. لازم به ذکر است که مدل PSR-VineCopula یک مدل چندکی است و بنابراین نتایج آن در بخش مدل‌های چندکی ارائه شده است.

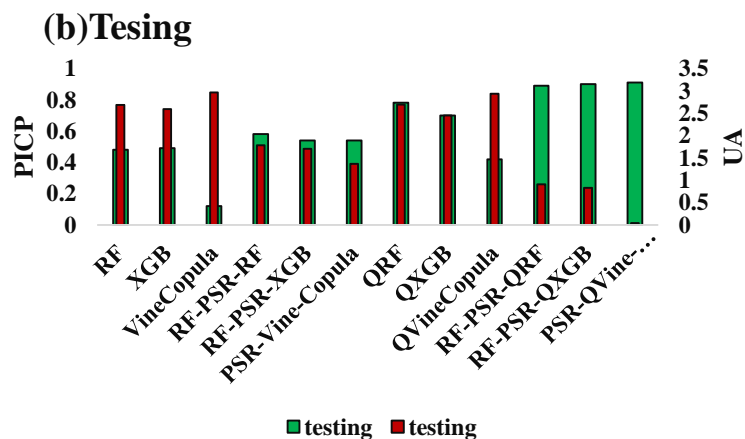
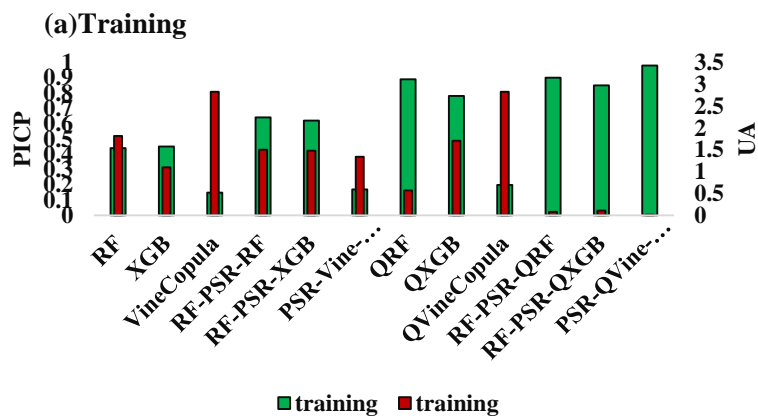
مدل‌های چندکی پایه (بدون PSR) نیز رفتار مشابهی با مدل‌های استاندارد دارند. هرچند مدل QRF در آموزش مقدار نسبتاً بالای $PICP=0.89$ را نشان می‌دهد، اما این مقدار در آزمون به ۰/۷۸ کاهش یافته و همزمان UA به ۲/۶۹ افزایش پیدا کرده است که نشان‌دهنده ناپایداری مدل‌های بدون PSR است. بهبود اصلی با اعمال PSR در مدل‌های چندکی رخ می‌دهد. در مدل‌های مبتنی بر PSR، مقادیر UB و LB در هر دو مرحله آموزش و آزمون به صفر نزدیک‌تر شده و دامنه عدم قطعیت به طور محسوسی کاهش یافته است. برای مثال، مقدار UA مدل PSR-QVineCopula در آزمون از ۲/۹۳ در حالت پایه به ۱/۳۷ کاهش یافته است. همزمان، شاخص پوشش به سطح بسیار مطلوبی رسیده است.

مدل‌های RF-PSR-QXGB و RF-PSR-QRF به پوشش‌های نسبتاً بالایی در آزمون ($PICP \approx 0.89-0.90$) دست یافته‌اند. مدل ترکیبی PSR-QVineCopula عملکردی استثنایی نشان می‌دهد. این مدل در آموزش به مقدار $PICP=0.98$ با دامنه عدم قطعیت بسیار کوچک ($UA=0.02$) دست یافته و در آزمون نیز مقدار بالای $PICP=0.91$ را همراه با کمترین UA در بین تمام مدل‌های چندکی ($UA=0.04$) ارائه کرده است. مقادیر UB و LB این مدل در آزمون ($UB=0.29$ و $LB=0.33$) نیز نشان‌دهنده آریب بسیار محدود و تمرکز پیش‌بینی‌ها حول مقادیر واقعی است.

جدول ۴- تحلیل عدم قطعیت مدل‌های مورد مطالعه

Models	Training				Testing			
	Without PSR							
	UB	LB	UA	PICP	UB	LB	UA	PICP

RF	۱.۶۷	-۰.۱۴	۱.۸۲	۰.۴۴	۰.۷۷	-۱.۹۱	۲.۶۸	۰.۴۸
XGB	۱.۳۰	۰.۱۹	۱.۱۰	۰.۴۵	۰.۶۹	-۱.۸۹	۲.۵۹	۰.۴۹
Vine copula	۱.۵۶	-۱.۲۶	۲.۸۳	۰.۱۵	۱.۰۶	-۱.۹۰	۲.۹۶	۰.۱۲
PSR								
RF-PSR-RF	۱.۴۶	-۰.۰۳	۱.۵۰	۰.۶۴	۰.۲۳	-۱.۵۵	۱.۷۸	۰.۵۸
RF-PSR-XGB	۱.۸۵	۰.۳۷	۱.۴۸	۰.۶۲	۰.۱۲۲	-۱.۵۸	۱.۷۰	۰.۵۴
PSR-VineCopula	۱.۳۶	-۰.۰۲۱	۱.۳۴	۰.۱۷	-۰.۰۴	-۱.۴۱	۱.۳۷	۰.۵۴
Without PSR								
Training				Testing				
	UB	LB	UA	PICP	UB	LB	UA	PICP
QRF	۰.۴۸	-۰.۰۸	۰.۵۷	۰.۸۹	۰.۷۸	-۱.۹۱	۲.۶۹	۰.۷۸
QXGB	۱.۰۷	-۰.۶۳	۱.۷۱	۰.۷۸	۰.۵۶	-۱.۸۸	۲.۴۵	۰.۷۰
QVine	۱.۵۶	-۱.۲۶	۲.۸۳	۰.۲	۱.۰۲	-۱.۹۰	۲.۹۳	۰.۴۲
PSR								
RF-PSR-QRF	۰.۲۶	۰.۱۷۶	۰.۰۸	۰.۹۰	۰.۳۰	۰.۶۱	۰.۹۱	۰.۸۹
RF-PSR-QXGB	۰.۲۷	۰.۱۶۰	۰.۱۱	۰.۸۵	۰.۰۷	۰.۷۵	۰.۸۳	۰.۹۰
PSR-QVineCopula	۰.۳۹۲	۰.۳۷۲	۰.۰۲	۰.۹۸	۰.۲۹	۰.۳۳	۰.۰۴	۰.۹۱



شکل ۶- اختلاف باندها و میزان PICP در مدل‌های در حالت با و بدون بازسازی فضای فاز

بحث و نتیجه‌گیری

تحلیل دینامیک سری زمانی شاخص خشکسالی (SPEI12) با استفاده از بازسازی فضای فاز (PSR)، مبنای بهبود مدل‌سازی در این پژوهش بود. یافته‌ها نشان داد که تأخیر زمانی بهینه (τ) بین ۵ تا ۱۰ ماه و بعد تعبیه بهینه (m) برابر با ۲ است. این نتایج مؤید دو ویژگی کلیدی سیستم اقلیمی منطقه است. اول وجود حافظه بلندمدت که اثرات شرایط اقلیمی را تا حدود یک سال حفظ می‌کند و دوم حضور یک ساختار دینامیکی غیرخطی با بعد مؤثر پایین که رفتار سیستم را از حالت تصادفی محض متمایز می‌سازد [۳۰، ۴۷]. این ساختار پیچیده، یکی از دلایل اصلی ناتوانی مدل‌های یادگیری ماشینی استاندارد در استخراج الگوهای تعمیم‌پذیر از داده‌های خام سری زمانی است. تکنیک PSR با «باز کردن» این دینامیک پنهان، یک بازنمایی معنادار از وضعیت سیستم ارائه می‌دهد که به مدل‌ها اجازه می‌دهد الگوهای واقعی را به‌جای نویز یاد بگیرند [۳۰].

انتخاب متغیرهای مؤثر در مدل‌سازی خشکسالی نشان‌داد که مهم‌ترین پیش‌بین‌ها، مقادیر گذشته شاخص SPEI به‌ویژه SPEI9 هستند. این امر تأکیدی بر ماندگاری زمانی و حافظه‌ی دینامیک خشکسالی دارد. همچنین، نقش دما و سرعت باد بیانگر ماهیت هیدرو-ترمودینامیکی خشکسالی منطقه‌ی ساحلی بندرعباس است که صرفاً ناشی از کمبود بارش نیست، بلکه با افزایش تبخیر و کاهش رطوبت سطحی تشدید می‌شود. تحلیل و این کاپولا نیز ساختار وابستگی‌های غیرخطی و نامتقارن بین متغیرها را آشکار کرده و علت ناتوانی مدل‌های خطی را در درک پویایی خشکسالی توضیح داده است [۳۶]. در نهایت، تلفیق خروجی و این کاپولا با مدل‌های یادگیری ماشینی، رویکردی نوآورانه برای لحاظ مستقیم روابط پیچیده در فرآیند آموزش ارائه می‌دهد [۱۹].

ارزیابی مدل‌ها نشان‌داد که بازسازی فضای فاز (PSR) نقش کلیدی در بهبود چشمگیر دقت پیش‌بینی خشکسالی دارد. در حالت بدون PSR،

حتی مدل‌های قدرتمندی مانند XGBoost و RF در مرحله آموزش و آزمون، عملکردی ضعیف با $R^2 \approx 0.19$ داشتند؛ زیرا به‌جای یادگیری الگوهای واقعی دینامیکی، تنها نویز داده‌ها را تقلید کردند. در مقابل، پس از اعمال PSR، کلیه مدل‌ها بهبود چشمگیری یافتند و مقدار $R^2 > 0.5$ در مرحله آزمون به‌دست آمد. این نشان می‌دهد که PSR با بازنمایی روشن از حالت سامانه در زمان، موجب شناسایی الگوهای معنادار و افزایش تعمیم‌پذیری مدل‌های یادگیری ماشینی می‌شود. افزون بر این، مدل ترکیبی PSR-Vine Copula توانست از طریق ترکیب بازسازی دینامیک سیستم و مدل‌سازی پیشرفته‌ی وابستگی غیرخطی، عملکردی فراتر از هر دو رویکرد به‌صورت منفرد ارائه دهد؛ بیانگر یک هم‌افزایی ساختاری میان تحلیل فاز و کاپولاها در مدل‌سازی خشکسالی‌های چندمقیاسی است [۳۴].

در این پژوهش، رویکرد پیش‌بینی احتمالی با استفاده از نسخه‌های چندکی مدل‌ها QRF، QXGBoost و QVineCopula و ارزیابی UNEEC به‌کار گرفته شد تا عدم قطعیت ذاتی خشکسالی به‌صورت واقع‌گرایانه مدل‌سازی شود. در حالت بدون PSR، مدل‌ها باندهای عدم قطعیت باریک و غیرواقعی تولید کردند که بیانگر اعتماد بیش‌ازحد (Overconfidence) و ناتوانی در برآورد پراکندگی داده‌ها بود؛ وضعیتی که در کاربرد عملی بسیار پرخطر است. با اعمال PSR، باندهای عدم قطعیت باریک‌تر و پایدارتر شکل گرفتند و شاخص پوشش بیش از ۰/۸ حاصل شد. این نتیجه نشان می‌دهد که مدل‌های چندکی، به‌ویژه PSR-QVineCopula، نه تنها دقت پیش‌بینی میانگین را افزایش دادند بلکه دامنه نوسانات آبی را نیز با صحت بالا تخمین زدند. چنین قابلیت‌هایی در مدیریت ریسک خشکسالی، تخصیص منابع آب و تصمیم‌گیری‌های کشاورزی بسیار حیاتی است و بخشی از چالش جعبه سیاه بودن مدل‌های ML را رفع می‌کند.

لازم به تأکید است که هدف این پژوهش در نسخه بازنگری‌شده، دستیابی به بیشینه‌سازی صرف

پیش‌بینی‌های پایدارتر را همراه با کمی‌سازی واقع‌گرایانه‌ی عدم قطعیت ارائه داد. سوم، شناسایی پیش‌بین‌های کلیدی نشان داد که پویایی خشکسالی در اقلیم بندرعباس تحت تأثیر هم‌زمان حافظه‌ی زمانی بلندمدت سیستم - تا حدود ده ماه گذشته - و مؤلفه‌های هیدروترم‌دینامیکی شامل دمای بیشینه و سرعت باد قرار دارد، که بیانگر ماهیت پیچیده و وابسته به انرژی این پدیده است.

این پژوهش نشان می‌دهد که آینده مدل‌سازی هیدرو-اقلیمی در گرو توسعه چارچوب‌های ترکیبی است که درک عمیق از دینامیک سیستم را با قدرت یادگیری الگوریتم‌های پیشرفته تلفیق می‌کنند. چارچوب توسعه‌یافته در این تحقیق می‌تواند به‌عنوان یک ابزار قدرتمند برای مدیران منابع آب و برنامه‌ریزان بخش کشاورزی در مناطق خشک و نیمه‌خشک جهت مدیریت ریسک و کاهش اثرات مخرب خشکسالی مورد استفاده قرار گیرد.

برای تحقیقات آینده پیشنهاد می‌شود که قابلیت تعمیم‌پذیری چارچوب ارائه‌شده در این پژوهش در مناطق اقلیمی مختلف ایران با ویژگی‌های دینامیکی متفاوت مورد ارزیابی قرار گیرد تا بتوان میزان سازگاری و پایداری عملکرد آن را در شرایط گوناگون اقلیمی سنجید. همچنین، گسترش مجموعه متغیرهای ورودی با بهره‌گیری از شاخص‌های دورپیوندی نظیر ENSO و NAO و داده‌های ماهواره‌ای شامل رطوبت خاک و پوشش گیاهی می‌تواند به درک جامع‌تری از محرک‌های خشکسالی منجر شود. افزون بر این، به‌کارگیری همین رویکرد ترکیبی در پیش‌بینی سایر پدیده‌های حدی اقلیمی مانند سیلاب‌ها و امواج گرما، گامی مؤثر در جهت توسعه مدل‌های هوشمند و قابل‌اعتماد برای مدیریت مخاطرات آب‌وهوایی خواهد بود.

شاخص‌های دقت نبوده است، بلکه تمرکز اصلی بر ارائه‌ی یک چارچوب ارزیابی واقع‌گرایانه، قابل اتکا و فیزیکی‌پذیر برای پیش‌بینی خشکسالی قرار گرفته است. در این راستا، استفاده از تفکیک کاملاً زمان‌محور داده‌ها و بازطراحی مجموعه متغیرهای ورودی، منجر به ارائه‌ی برآوردهایی شده است که هرچند محافظه‌کارانه‌تر هستند، اما قابلیت تعمیم‌پذیری بالاتری در شرایط عملی دارند.

از این منظر، مقادیر دقت گزارش‌شده در این پژوهش نه به‌عنوان هدف نهایی، بلکه به‌عنوان پیامد طبیعی یک چارچوب مدل‌سازی منسجم مبتنی بر دینامیک سیستم، اعتبارسنجی زمان‌محور و تفسیر فیزیکی نتایج تلقی می‌شوند. محدودیت‌های مرتبط با ساختار داده‌ها، حافظه زمانی سیستم و عدم قطعیت ذاتی فرآیند خشکسالی نیز به‌صورت شفاف در این بخش مورد بحث قرار گرفته‌اند.

این پژوهش با هدف توسعه یک چارچوب ترکیبی پیشرفته برای بهبود دقت و ارزیابی عدم قطعیت در پیش‌بینی خشکسالی بلندمدت (SPEI12) در منطقه فراخشک و ساحلی بندرعباس انجام شد. نتایج به‌وضوح نشان داد که رویکردهای استاندارد یادگیری ماشین به‌دلیل ناتوانی در درک دینامیک غیرخطی و حافظه بلندمدت سیستم اقلیمی، عملکرد قابل قبولی ندارند.

این پژوهش با سه دستاورد بنیادی همراه بود. نخست، نقش حیاتی بازسازی فضای فاز (PSR) در بهبود چشمگیر عملکرد مدل‌های یادگیری ماشین در پیش‌بینی خشکسالی اثبات شد، به‌گونه‌ای که دقت مدل‌ها افزایش یافت و رفتار تعمیم‌پذیرتری در مرحله آزمون زمان‌محور از خود نشان دادند. دوم، چارچوب ترکیبی نوآورانه‌ی PSR-QVineCopula توسعه یافت که با ادغام بازسازی دینامیکی، مدل‌سازی وابستگی‌های غیرخطی و رگرسیون چندکی،

References

- [1] Abdallah, M., Mohammadi, B., Zaroug, M. A., Omer, A., Cheraghalizadeh, M., Eldow, M. E., & Duan, Z. (2022). Reference

evapotranspiration estimation in hyper-arid regions via D-vine copula based-quantile regression and comparison with empirical approaches and machine learning models. *Journal of Hydrology*:

- Regional Studies*, 44, 101259. <https://doi.org/10.1016/j.ejrh.2022.101259>.
- [2] Abdallah, M., Zhang, K., Chao, L., Omer, A., Hassaballah, K., Welde Reda, K., ... & Nour, O. M. (2024). A D-vine copula-based quantile regression towards merging satellite precipitation products over rugged topography: a case study in the upper Tekeze–Atbara Basin. *Hydrology and Earth System Sciences*, 28(5), 1147-1172. <https://doi.org/10.5194/hess-28-1147-2024>.
- [3] Aas, K., Czado, C., Frigessi, A., & Bakken, H. (2009). Pair-copula constructions of multiple dependence. *Insurance: Mathematics and economics*, 44(2), 182-198. <https://doi.org/10.1016/j.insmatheco.2007.02.001>.
- [4] Breiman, L. (2001). Random forests. *Machine learning*, 45(1), 5-32. <https://doi.org/10.58496/BJML/2024/007>.
- [5] Chen, C., Liaw, A., & Breiman, L. (2004). *Using random forest to learn imbalanced data*. Statistics Department of University of California at Berkeley. Berkeley. Technical Report 666.
- [6] Cannon, A. J. (2011). Quantile regression neural networks: Implementation in R and application to precipitation downscaling. *Computers & geosciences*, 37(9), 1277-1284. <https://doi.org/10.1016/j.cageo.2010.07.005>.
- [7] Dikshit, A., Pradhan, B., & Alamri, A. M. (2020). Short-term spatio-temporal drought forecasting using random forests model at New South Wales, Australia. *Applied Sciences*, 10(12), 4254. <https://doi.org/10.3390/app10124254>.
- [8] Danandeh Mehr, A., Rikhtehgar Ghiasi, A., Yaseen, Z. M., Sorman, A. U., & Abualigah, L. (2023). A novel intelligent deep learning predictive model for meteorological drought forecasting. *Journal of Ambient Intelligence and Humanized Computing*, 14(8), 10441-10455. <https://doi.org/10.1007/s12652-022-03701-7>.
- [9] Elidan, G. (2010). Copula bayesian networks. *Advances in neural information processing systems*, 23.
- [10] Feng, P., Wang, B., Li Liu, D., & Yu, Q. (2019). Machine learning-based integration of remotely-sensed drought factors can improve the estimation of agricultural drought in South-Eastern Australia. *Agricultural Systems*, 173, 303-316. <https://doi.org/10.1016/j.agsy.2019.03.015>.
- [11] Fraser, A. M., & Swinney, H. L. (1986). Independent coordinates for strange attractors from mutual information. *Physical review A*, 33(2), 1134. <https://doi.org/10.1109/18.32121>.
- [12] Goldblatt, R., You, W., Hanson, G., & Khandelwal, A. K. (2016). Detecting the boundaries of urban areas in india: A dataset for pixel-based image classification in google earth engine. *Remote Sensing*, 8(8), 634. <https://doi.org/10.3390/rs8080634>.
- [13] Guo, Q., He, Z., & Wang, Z. (2024). Monthly climate prediction using deep convolutional neural network and long short-term memory. *Scientific Reports*, 14(1), 17748. <https://doi.org/10.1038/s41598-024-68906-6>.
- [14] Ghafari, L., & Parvishi, A. (2025). Climate projection and drought assessment in the lake Urmia basin using LSTM-based downscaling of GCM models under SSP scenarios. *Physics and Chemistry of the Earth, Parts A/B/C*, 104134. <https://doi.org/10.1016/j.pce.2025.104134>.
- [15] Huber, P. J. (1973). Robust regression: asymptotics, conjectures and Monte Carlo. *The annals of statistics*, 799-821. <https://doi.org/10.1214/aos/1176342503>.
- [16] Bergmeir, C., Hyndman, R. J., & Koo, B. (2018). A note on the validity of cross-validation for evaluating autoregressive time series prediction. *Computational Statistics & Data Analysis*, 120, 70-83.

- [17] Joe, H. (1996). Families of m-variate distributions with given margins and m (m-1)/2 bivariate dependence parameters. *Lecture notes-monograph series*, 120-141. <https://doi.org/10.1214/lnms/1215452614>.
- [18] Javadi, A., Ghahremanzadeh, M., Sassi, M., Javanbakht, O., & Hayati, B. (2024). Impact of climate variables change on the yield of wheat and rice crops in Iran (application of stochastic model based on Monte Carlo simulation). *Computational Economics*, 63(3), 983-1000. <https://doi.org/10.1007/s10614-023-10389-0>.
- [19] Konstantelos, I., Sun, M., Tindemans, S. H., Issad, S., Panciatici, P., & Strbac, G. (2018). Using vine copulas to generate representative system states for machine learning. *IEEE Transactions on Power Systems*, 34(1), 225-235. <https://doi.org/10.1109/TPWRS.2018.2859367>.
- [20] Kursu, M. B., & Rudnicki, W. R. (2010). Feature selection with the Boruta package. *Journal of statistical software*, 36, 1-13. <https://doi.org/10.18637/jss.v036.i11>.
- [21] Koenker, R., & Bassett Jr, G. (1978). Regression quantiles. *Econometrica: journal of the Econometric Society*, 33-50. <https://doi.org/10.1080/03610910701723963>.
- [22] Kheyruri, Y., Sharafati, A., & Neshat, A. (2023). The socioeconomic impact of severe droughts on agricultural lands over different provinces of Iran. *Agricultural Water Management*, 289, 108550. <https://doi.org/10.1016/j.agwat.2023.108550>.
- [23] Meinshausen, N., & Ridgeway, G. (2006). Quantile regression forests. *Journal of machine learning research*, 7(6).
- [24] Muñoz-Sabater, J., Dutra, E., Agustí-Panareda, A., Albergel, C., Arduini, G., Balsamo, G., ... & Thépaut, J. N. (2021). ERA5-Land: A state-of-the-art global reanalysis dataset for land applications. *Earth system science data*, 13(9), 4349-4383. <https://doi.org/10.5194/essd-2021-82>
- [25] Nazeri Tahroudi, M., Ramezani, Y., de Michele, C., & Mirabbasi, R. (2022). Multivariate analysis of rainfall and its deficiency signatures using vine copulas. *International Journal of Climatology*, 42(4), 2005-2018. <https://doi.org/10.1002/joc.7349>.
- [26] Ni, L., Wang, D., Wu, J., Wang, Y., Tao, Y., Zhang, J., ... & Xie, F. (2020). Vine copula selection using mutual information for hydrological dependence modeling. *Environmental research*, 186, 109604. <https://doi.org/10.1016/j.envres.2020.109604>.
- [27] Neetu, & Ray, S. S. (2019). Exploring machine learning classification algorithms for crop classification using Sentinel 2 data. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 42, 573-578. <https://doi.org/10.5194/isprs-archives-XLII-3-W6-573-2019>.
- [28] Rahmati, O., Falah, F., Dayal, K. S., Deo, R. C., Mohammadi, F., Biggs, T., ... & Bui, D. T. (2020). Machine learning approaches for spatial modeling of agricultural droughts in the south-east region of Queensland Australia. *Science of the total environment*, 699, 134230. <https://doi.org/10.1016/j.scitotenv.2019.134230>.
- [29] Sklar, M. (1959). Fonctions de répartition à n dimensions et leurs marges. In *Annales de l'ISUP* (Vol. 8, No. 3, pp. 229-231).
- [30] Sivakumar, B. (2002). A phase-space reconstruction approach to prediction of suspended sediment concentration in rivers. *Journal of Hydrology*, 258(1-4), 149-162. [https://doi.org/10.1016/S0022-1694\(01\)00573-X](https://doi.org/10.1016/S0022-1694(01)00573-X).
- [31] Samantaray, S., & Ghose, D. K. (2022). Prediction of S12-MKII rainfall simulator experimental runoff data sets using hybrid PSR-SVM-FFA approaches. *Journal of Water and Climate Change*,

- 13(2), 707-734. <https://doi.org/10.2166/wcc.2021.221>
- [32] Sahin, E. K. (2020). Assessing the predictive capability of ensemble tree methods for landslide susceptibility mapping using XGBoost, gradient boosting machine, and random forest. *SN Applied Sciences*, 2(7), 1308. <https://doi.org/10.1007/s42452-020-3060-1>.
- [33] Takens, F. (1981). Detecting strange attractors in turbulence. In D. Rand & L.-S. Young (Eds.), *Dynamical Systems and Turbulence, Lecture Notes in Mathematics* (Vol. 898, pp. 366–381). Springer. <https://doi.org/10.1007/BFb0091924>.
- [34] Tavosi, M., Vafakhah, M., Sadeghi, S. H., Shekohideh, H., & Moosavi, V. (2025). Comparative assessment of Watershed Hydrological Health (WHH) using multi-criteria decision-making approach based on PSR framework. *Journal of Environmental Management*, 373, 123833. <https://doi.org/10.1016/j.jenvman.2024.123833>.
- [35] Tayebikhorami, A., Mahdavi, R & nohegar, A. (2015). Survey of Consumption water resources priorities in normal conditions and drought (case study: Bandar Abbas city). *Iranian Water Research Journal*. 13(1), 29-39 [in Farsi].
- [36] Wang, S., Liu, Y., Wang, W., Zhao, G., & Liang, H. (2024). Interpretable machine learning guided by physical mechanisms reveals drivers of runoff under dynamic land use changes. *Journal of Environmental Management*, 367, 121978. <https://doi.org/10.1016/j.jenvman.2024.121978>.
- [37] Wu, T., Bai, J., & Han, H. (2022). Short-term agricultural drought prediction based on D-vine copula quantile regression in snow-free unfrozen surface area, China. *Geocarto International*, 37(25), 9320-9338. <https://doi.org/10.1080/10106049.2021.2017015>.
- [38] Wang, M., Jiang, S., Ren, L., Xu, C. Y., Wei, L., Cui, H., ... & Yang, X. (2022). The development of a nonstationary standardised streamflow index using climate and reservoir indices as covariates. *Water Resources Management*, 36(4), 1377-1392. <https://doi.org/10.1007/s11269-022-03088-2>.
- [39] 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*, 9, 1679. <https://doi.org/10.3389/fpsyg.2018.01679>.
- [40] Van Dijk, M., Morley, T., Rau, M. L., & Saghari, Y. (2021). A meta-analysis of projected global food demand and population at risk of hunger for the period 2010–2050. *Nature food*, 2(7), 494-501. <https://doi.org/10.1038/s43016-021-00322-9>.
- [41] Zare, S., Abtahi, A., Dehghani, M., Shamsi, S. R. F., Baghernejad, M., & Lagacherie, P. (2024). Quantile random forest technique for soil moisture contents digital mapping, Sarvestan Plain, Iran. *In Advanced Tools for Studying Soil Erosion Processes* (pp. 351-368). Elsevier. <https://doi.org/10.1016/B978-0-443-22262-7.00001-1>.
- [42] Zegaar, A., Ounoki, S., & Telli, A. (2024). Machine learning for groundwater quality classification: A step towards economic and sustainable groundwater quality assessment process. *Water Resources Management*, 38(2), 621-637. <https://doi.org/10.1007/s11269-023-03690-y>.
- [43] Zamani, H., Pakdaman, Z., Shakari, M., Bazrafshan, O., & Jamshidi, S. (2025). Enhancing drought monitoring with a multivariate hydrometeorological index and machine learning-based prediction in the south of Iran. *Environmental Science and Pollution Research*, 32(9), 5605-5627.

Application of Phase Space Reconstruction in Enhancing the Performance of Machine Learning Models for Drought Prediction (*Case Study: Bandar Abbas Synoptic Station*) (Research Paper)

1- Mehrnaz Yahyizadeh, PhD Student, Department of Natural Resources Engineering, Faculty of Agricultural and Natural Resources Engineering, University of Hormozgan, Bandar Abbas, Iran.

2- Umm El-Bazin Bazrafshan*, Professor, Department of Natural Resources Engineering, Faculty of Agricultural and Natural Resources Engineering, University of Hormozgan, Bandar Abbas, Iran.

O.bazrafshan@hormozgan.ac.ir

3- Nawazollah Moradi, Assistant Professor, Department of Natural Resources Engineering, Faculty of Agricultural and Natural Resources Engineering, University of Hormozgan, Bandar Abbas, Iran.

4- Hossein Zamani, Assistant Professor, Department of Mathematics and Statistics, Faculty of Basic Sciences, University of Hormozgan, Bandar Abbas, Iran.

5- Mehdi Biniaz, Assistant Professor, Department of Natural Resources Engineering, Faculty of Agricultural and Natural Resources Engineering, University of Hormozgan, Bandar Abbas, Iran.

Received: 29 Nov. 2025

Accepted: 1 Feb. 2026

Abstract

Drought, as one of the most complex and multidimensional climatic phenomena, is inherently associated with high uncertainty and nonlinear dependencies among climatic variables. In this study, a hybrid framework integrating phase space reconstruction (PSR), Vine Copula, and quantile regression was developed to improve the accuracy and interpretability of drought prediction in the hyper-arid climate of Bandar Abbas over the period 1960–2022. The results demonstrate that phase space reconstruction, by unveiling the hidden dynamical structure of the SPEI time series, significantly enhances model performance, such that the coefficient of determination (R^2) in the testing phase increases from approximately 0.10 to more than 0.80. The Vine Copula approach effectively captures the nonlinear and asymmetric dependencies among climatic variables and, when combined with PSR, provides a coherent framework for representing the dynamic behavior of the drought system. Quantile-based models, including QRF, QXGBoost, and QVineCopula, not only improve predictive accuracy but also enable data-driven quantification of uncertainty, resulting in superior coverage of prediction intervals (PICP = 0.91). Variable importance analysis highlights the dominant role of lagged SPEI components—particularly SPEI9 and SPEI8—along with maximum temperature and wind speed, in governing regional drought dynamics. Overall, the hybrid PSR–QVineCopula model achieves an optimal balance between accuracy, robustness, and interpretability, and offers a novel framework for both deterministic and probabilistic drought prediction in hyper-arid regions.

Keywords: Drought, Phase Space Reconstruction (PSR), Vine Copula, Quantile Regression, Probabilistic Prediction, Uncertainty.