

بررسی عملکرد الگوریتم‌های Kstar و GPR در مدل‌سازی شاخص خشکسالی هواشناسی RDI (مطالعه موردی: شرق حوضه دریاچه ارومیه)

مرضیه صادقیان آقکندی^۱، حسین رضایی*^۲، کیوان خلیلی^۳، فرشاد احمدی^۴

^۱ دانشجوی دکتری منابع آب، گروه علوم و مهندسی آب، دانشکده کشاورزی، دانشگاه ارومیه

^۲ استاد گروه علوم و مهندسی آب، دانشکده کشاورزی، دانشگاه ارومیه

^۳ دانشیار گروه علوم و مهندسی آب، دانشکده کشاورزی، دانشگاه ارومیه

^۴ استادیار گروه هیدرولوژی و منابع آب، دانشکده مهندسی آب، دانشگاه شهید چمران اهواز

دریافت: ۱۴۰۱/۱۱/۲۴، بازنگری: ۱۴۰۲/۱/۲۴، پذیرش: ۱۴۰۲/۲/۲۷، نشر آنلاین: ۱۴۰۲/۲/۲۷

چکیده

خشکسالی یکی از بلاهای طبیعی است که سالانه خسارات فراوانی را برای جوامع مختلف به دنبال دارد. پایش خشکسالی‌ها می‌تواند اطلاعات بالارزشی جهت تدوین برنامه‌های مقابله با خشکسالی و کاهش خسارات مربوط به آن در اختیار مدیران و برنامه‌ریزان منابع آب بگذارد. در تحقیق حاضر، خشکسالی هواشناسی چهار ایستگاه سینوپتیک (Synoptic Scale Meteorology) اهر، جلفا، تبریز و مراغه در استان آذربایجان شرقی، شرق دریاچه ارومیه با استفاده از شاخص خشکسالی RDI در دوره آماری ۱۹۵۵ تا ۲۰۱۹ مورد بررسی و تحلیل قرار گرفت. برای محاسبه شاخص RDI از داده‌های بارش و تبخیر تعرق پتانسیل ماهانه با روش فائو پنمن مانیت (FP-M) استفاده شد. شاخص RDI برای هر ایستگاه به صورت جداگانه و در سه مقیاس زمانی ۶، ۹ و ۱۲ ماهه محاسبه گردید و نتیجه محاسبات نشان داد در مقیاس زمانی ۱۲ ماهه پایش خشکسالی بهتر از مقیاس‌های زمانی دیگر می‌باشد. سپس با استفاده از الگوریتم‌های Kstar و GPR شاخص RDI در سه مقیاس زمانی مدل‌سازی شد. نتایج حاصل از معیارهای ارزیابی، عملکرد بالای الگوریتم‌های ذکر شده را نشان داد به طوری که مقدار عددی ضریب همبستگی الگوریتم GPR برای همه ایستگاه‌ها ۰/۹۲ و همین ضریب با الگوریتم Kstar در بازه ۰/۹۱ تا ۰/۹۲ قرار دارد. همچنین مقدار عددی RMSE با الگوریتم GPR بازه ۰/۳۱ تا ۰/۳۹ و الگوریتم Kstar ۰/۳۲ تا ۰/۵۱ را به دست آورد. از این نتایج می‌توان ادعا کرد که الگوریتم GPR با دقت بالاتری شاخص خشکسالی RDI را مدل کرده است.

کلیدواژه‌ها: الگوریتم Kstar، پایش خشکسالی RDI، رگرسیون فرایند گاوسی، شرق دریاچه ارومیه.

۱- مقدمه

(Mishra و Singh، ۲۰۱۰؛ Beyaztas و Yaseen، ۲۰۱۹). در یک طبقه‌بندی کلی، رویدادهای خشکسالی به چهار دسته مختلف، یعنی خشکسالی‌های هواشناسی، هیدرولوژیکی، کشاورزی و اجتماعی-اقتصادی طبقه‌بندی می‌شوند (Wilhite و Glantz، ۱۹۸۵؛ Khadr، ۲۰۱۶). برخی از شاخص‌های خشکسالی هواشناسی شناخته شده و متداول در پایش خشکسالی‌ها شامل شاخص شدت خشکسالی

خشکسالی یک رویداد شدید هیدرولوژیکی است که می‌تواند مشکلات جدی در زندگی انسان ایجاد کند. در این زمینه، می‌تواند اثرات نامطلوبی بر تأمین و کیفیت آب، سلامت عمومی، بهره‌وری کشاورزی، تخریب زمین، بیابان‌زایی، قحطی و ... بگذارد (Moradkhani و Li، ۲۰۱۴؛ Li و همکاران، ۲۰۲۰). خشکسالی به عنوان یکی از بلاهای طبیعی توسط برخی عناصر از جمله افزایش تقاضای آب به دلیل افزایش جمعیت و همچنین فرآیندهای جوی و هیدرولوژیکی تشدید می‌شود.



همان ساختار SPI به علت بهره‌گیری از تبخیر و تعرق پتانسیل در ساختار خود برای ارزیابی خشکسالی کشاورزی و هیدرولوژیکی از قابلیت‌های ویژه‌ای برخوردار می‌باشد (شکوهی، ۱۳۹۲). در منطقه مورد مطالعه تاکنون مقایسه عملکرد دو مدل Kstar و GPR بر روی شاخص RDI انجام نشده است. لذا با توجه به قابلیت این دو مدل، در این پژوهش پایش خشکسالی هواشناسی در حوضه شرق دریاچه ارومیه در یک دوره ۶۵ ساله، بررسی و تحلیل نتایج حاصل از این دو مدل در شاخص RDI مورد توجه قرار گرفت.

۲- روش تحقیق

۱-۲- داده‌ها و منطقه مورد مطالعه

شرق حوضه دریاچه ارومیه با وسعتی به میزان ۱۰۸۴,۳۸ کیلومتر مربع بین عرض شمالی حداقل ۳۶ و حداکثر ۳۸ درجه و طول شرقی حداقل ۴۶ و حداکثر ۴۸ درجه در شرقی‌ترین قسمت حوضه آبریز دریاچه ارومیه واقع شده است. در این محدوده جریانات ورودی عبارت‌اند از: جریانات سطحی، بارش‌های باران و برف و مسیل‌های دائمی. میانگین بارش سالانه در دشت‌های این محدوده ۳۳۰ میلی‌متر محاسبه شده است. اطلاعات هواشناسی مورد استفاده در این پژوهش از داده‌های ماهانه بارش و تبخیر-تعرق پتانسیل ایستگاه‌های سینوپتیک اهر، تبریز، جلفا و مراغه در سال‌های ۱۹۵۵ تا ۲۰۱۹ برای برآورد شاخص RDI استفاده شده است. در شکل (۱) موقعیت جغرافیایی محدوده مطالعاتی و موقعیت ایستگاه‌های هواشناسی حوضه شرق دریاچه ارومیه نشان داده شده است.

۲-۲- شاخص شناسایی خشکسالی RDI

این شاخص براساس دو فراسنج بارش و تبخیر-تعرق پتانسیل عمل می‌کند. نحوه محاسبه شاخص به شرح زیر است (Tsakiris و Vangelis, ۲۰۰۷). ابتدا با استفاده از رابطه (۱) برای هر سال i از دوره آماری مورد مطالعه مقداری با عنوان α_0 محاسبه می‌شود:

$$\alpha_0^i = \frac{\sum_{j=1}^{12} P_{ij}}{\sum_{j=1}^{12} ET_{ij}} \quad (1)$$

P بارندگی، ET تبخیر-تعرق پتانسیل در زامین ماه سال i . مقدار i از یک تا N (تعداد سال‌هایی است که آمار آن در دسترس است) تغییر خواهد کرد. جهت محاسبه شاخص معیار شده RDI (RDI_s) از مقادیر α_0 لگاریتم گرفته که عددی با عنوان y_i به دست می‌آید. سپس میانگین حسابی و انحراف معیار این اعداد محاسبه و به ترتیب \bar{y}_k و σ_{y_k} نامیده می‌شوند و نهایتاً شاخص معیار RDI به دست می‌آید.

پالمر (PDSI)^۱، شاخص شناسایی خشکسالی (RDI)^۲، شاخص بارش استاندارد شده (SPI)^۳ و شاخص استاندارد تبخیر و تعرق بارش است (SPEI)^۴ (Palmer, ۱۹۶۵؛ Tsakiris و همکاران، ۲۰۰۷؛ Mckee و همکاران، ۱۹۹۳؛ Vicente-Serrano و همکاران، ۲۰۱۰).

با توجه به نقش قابل توجه تبخیر و تعرق در تعادل آب، لازم است اثر آن در هنگام مطالعه خشکسالی در یک منطقه خاص مورد توجه قرار گیرد. بر این اساس، RDI که هم بارش و هم تبخیر و تعرق را در بر می‌گیرد، می‌تواند به‌عنوان یک شاخص قابل اعتماد برای پایش خشکسالی در نظر گرفته شود (معینی‌فر و همکاران، ۱۴۰۰).

با توجه به اهمیت خشکسالی به‌عنوان یک پدیده طبیعی در مطالعات هیدرولوژیکی و هواشناسی، پایش و پیش‌بینی آن با رویکردی مناسب می‌تواند حائز اهمیت باشد. پیش‌بینی خشکسالی نقش اساسی در مدیریت ریسک، کاهش اثرات خشکسالی بر منابع آب موجود و استفاده بهینه از آن‌ها، امکان تصمیم‌گیری منطقی توسط تصمیم‌گیرندگان برای به حداقل رساندن خسارات ناشی از خشکسالی و همچنین برنامه‌ریزی و مدیریت پروژه‌های منابع آب دارد (Khadr و همکاران، ۲۰۱۶؛ Madrigal و همکاران، ۲۰۱۸؛ Beyaztas و Yaseen, ۲۰۱۹).

در میان مدل‌های کارشده تا به امروز، مدل‌های منفرد Kstar^۵ و GPR^۶ جدیدترین مدل برای پیش‌بینی خشکسالی می‌باشند. تاکنون شاخص‌های متعددی توسط محققان مختلف برای بررسی خشکسالی در دنیا معرفی و استفاده شده‌اند. شاخص‌های پالمر، ذخیره آب سطحی، درصد نرمال، دهک‌ها، بارش استاندارد، رطوبت محصول، رطوبت خاک، محصول ویژه، بارش سراسری، ناهنجاری بارش، خشکسالی احیایی و بارش مؤثر از جمله این شاخص‌ها هستند که توسط افراد مختلف ارائه شده‌اند (فاطمی و همکاران، ۱۳۹۶).

در سالیان اخیر، شاخص‌های جدیدتری نظیر شاخص شناسایی خشکسالی یا RDI استفاده شده است. شاخص RDI در ایران و جهان سابقه طولانی ندارد و ارائه آن مربوط به سال ۲۰۰۴ می‌باشد (معینی‌فر و همکاران، ۱۴۰۰). یکی دیگر از شاخص‌های پرکاربرد SPI می‌باشد که توسط Mckee و همکارانش (۱۹۹۳) ارائه شده است (میرعباسی نجف‌آبادی و همکاران، ۱۳۹۶).

شاخص RDI به دلیل در نظر گرفتن توأم بارش و تبخیر و تعرق پتانسیل، با استقبال بیشتری نسبت به سایر شاخص‌ها روبرو بوده و در بسیاری از مطالعات مبنای مقایسه شدت خشکسالی و برآورد خسارات بالقوه آن در سطح ملی قرار گرفته است. شاخص RDI با

4. Standardized Precipitation Evapotranspiration Index
5. Kstar Algorithm
6. Gaussian Process Regression

1. Palmer Drought Severity Index
2. Reconnaissance Drought Index
3. Standardized Precipitation Index

دیگر توصیف شود (Granata و همکاران، ۲۰۱۹). همان گونه که ذکر شد رگرسیون Kstar، از یک تابع فاصله آنتروپی استفاده می-کند. فاصله آنتروپیک برای به دست آوردن نمونه‌هایی که بیشترین شباهت را از مجموعه داده‌ها به یکدیگر دارند استفاده می‌شود. در نظر بگیرد که a و b نمونه‌های مورد بررسی باشد در این صورت P^* را می‌توان به عنوان احتمال هر مسیر توصیف کرد که از a به b است. بنابراین رابطه P به شرح زیر قابل بیان می‌باشد (Ekmekcioğlu و همکاران، ۲۰۲۰):

$$P^*(b|a) = \sum_{t \in p: t(a)=b}^N p(t) \quad (3)$$

که در آن t نشان‌دهنده مقدار T (یک مجموعه‌ای از تبدیلات داده‌هاست) بوده و p یک تابع احتمال است. با در نظر گرفتن این که P^* شرایط زیر را دارد:

$$\sum_b^N P^*(b|a) = 1 \quad (4)$$

$$0 \leq P^*(b|a) \leq 1$$

با توجه به روابط فوق تابع Kstar به شرح زیر بیان می‌شود:

$$K^*(b|a) = -\log P^*(b|a) \quad (5)$$

رابطه (۵) دارای ویژگی‌های زیر است:

$$K^*(b|a) \geq 0$$

$$K^*(b|a) + K^*(c|b) \geq K^*(c|a)$$

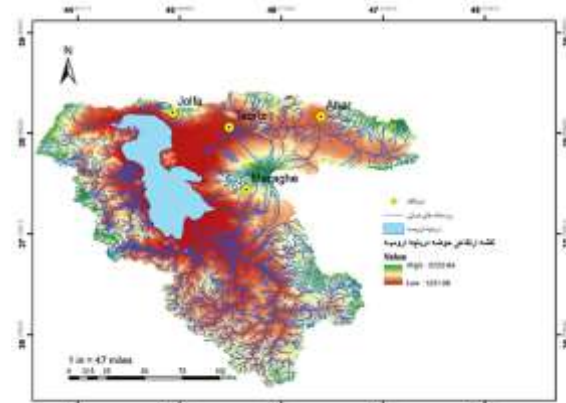
روابط فوق، نماینده اعداد صحیح بوده و برای اعداد پیوسته به صورت زیر بازنویسی می‌شوند:

$$P^*(b|a) = P^*(i) = \frac{s}{\sqrt{2s-s^2}} \left(\frac{1-\sqrt{2s-s^2}}{1-s} \right) \quad (6)$$

where $i = |a-b|$

$$K^*(b|a) = K^*(i) = \frac{1}{2} \log(2s-s^2) - \log(s) + i \left[\log(1-s) - \log(1-\sqrt{2s-s^2}) \right] \quad (7)$$

که در آن s پارامتر مدل بوده و بین صفر و یک متغیر است. با در دست داشتن این روابط می‌توان با استفاده از مقادیر احتمالاتی محاسبه شده مناسب‌ترین نمونه را برای داده مورد نظر انتخاب نمود.



شکل ۱- موقعیت جغرافیایی ایستگاه‌های هواشناسی شرق حوضه دریاچه ارومیه

$$RDI_{st(k)}^{(i)} = \frac{y_k^{(i)} - \bar{y}_k}{\hat{\sigma}_{y_k}} \quad (2)$$

این شاخص می‌تواند با در نظر گرفتن مقیاس‌های مختلف زمانی و به صورت سالانه وضعیت خشکسالی را پایش کند. تفسیر مقادیر به دست آمده از این روش بر اساس جدول (۱) انجام می‌گیرد (Vangelis و Tsakiris، ۲۰۰۷).

جدول ۱- طبقه‌بندی شاخص RDI

مقدار شاخص RDI	طبقه‌بندی خشکسالی
بیشتر از ۲	به شدت مرطوب
۱/۵ تا ۱/۹۹	خیلی مرطوب
۱ تا ۱/۴۹	نسبتاً مرطوب
۰/۹۹ تا -۰/۹۹	متوسط
-۱ تا -۱/۴۹	نسبتاً خشک
-۱/۵ تا -۱/۹۹	خیلی خشک
کمتر از -۲	به شدت خشک

۲-۳- الگوریتم Kstar^v

الگوریتم Kstar را می‌توان به عنوان روشی برای تجزیه و تحلیل خوشه‌ای تعریف کرد که هدف اصلی آن تقسیم n مشاهده به K خوشه است به طوری که در آن، هر مشاهده به خوشه با نزدیک‌ترین میانگین تعلق دارد (Ekmekcioğlu و همکاران، ۲۰۲۰). الگوریتم Kstar را می‌توان به عنوان یک روش یادگیری مبتنی بر نمونه توصیف کرد که از تئوری آنتروپی به عنوان اندازه-گیری فاصله استفاده می‌کند. این روش با ارائه یک رویکرد سازگار برای مدیریت خصوصیات ارزشمند واقعی، ویژگی‌های نمادین و مقادیر از دست‌رفته امکان دریافت اطلاعات را از داده‌ها موجود به حداکثر می‌رساند. در الگوریتم Kstar فاصله از یک نمونه به نمونه دیگر می‌تواند به عنوان پیچیدگی تبدیل یک نمونه به نمونه

۲-۴- رگرسیون فرآیند گاوسی (GPR)^۸

در نظریه آمار و احتمال یک فرایند گاوسی یک مدل آماری است که در آن مشاهدات در دامنه پیوسته رخ می‌دهد. در واقع فرآیند گاوسی، یک فرآیند تصادفی است که به صورت دنباله‌ای از متغیرهای تصادفی شناخته می‌شود و در آن متغیرهای تصادفی معمولاً برحسب؛ به طور مثال زمان مرتب شده‌اند و هر زیر مجموعه متناهی از این متغیرها دارای توزیع گاوسی (نرمال) چندمتغیره هستند. در یک فرایند گاوسی هر نقطه از فضای ورودی یک متغیر تصادفی با توزیع نرمال است. علاوه بر این هر مجموعه متناهی از این متغیرهای تصادفی دارای توزیع گاوسی چند متغیره است. توزیع فرایند گاوسی توزیع مشترک از تمام این متغیرهای تصادفی (شمارا و نامحدود) است. از دید یک الگوریتم یادگیری ماشین، یک فرایند گاوسی، اندازه‌گیری شباهت بین نقاط (همان تابع کرنل) برای پیش‌بینی نقاط جدید از داده‌های آموزشی است. فرآیند گاوسی مجموعه‌ای از متغیرهای تصادفی است که تعداد دلخواه محدودی از آن‌ها با توزیع‌های گاوسی ادغام شده‌اند. فرایند گاوسی به طور کامل توسط تابع میانگین $m(x)$ و تابع کوواریانس $k(x, x')$ آن تعیین می‌شود. این فرایند، تعمیمی طبیعی از توزیع گاوسی است که میانگین و کوواریانس آن به ترتیب یک بردار و یک ماتریس است. مدل‌های رگرسیون فرایند گاوسی بر مبنای این فرض هستند که مشاهدات تنظیم باید حامل اطلاعاتی درباره همدیگر باشد. یکی از ویژگی‌های مهم در فرایند گاوسی وجود توابع کوواریانس متنوع در آن است که به محقق این امکان را می‌دهد که از میان آن‌ها گزینه‌ای مناسبی انجام دهد. این مدل‌ها می‌توانند توزیع‌هایی را بین توابعی با تعداد یک یا چندین متغیر ورودی مشخص کنند هنگامی که چنین تابعی میانگین پاسخ را در یک مدل رگرسیونی با خطاهای گاوسی تعریف می‌کند، می‌توان از محاسبات ماتریس برای استنتاج بهره‌گیری کرد؛ این مسئله برای مجموعه داده‌هایی با بیش از هزار نمونه امکان‌پذیر است. فرآیندهای گاوسی در مدل‌سازی آماری بسیار پُراهمیت‌اند، زیرا خصوصیات نرمالی دارند (Neal, ۱۹۹۷). مجموعه داده S با n مشاهده را در نظر بگیرید $\mathcal{S} = \{(x_i, y_i) | i = 1, \dots, n\}$ که در آن x_i بردار ورودی با D بعد و y_i خروجی اسکالر یا هدف می‌باشد. این مجموعه متشکل از دو جزء ورودی و خروجی به‌عنوان نقاط نمونه یا تجربی معرفی خواهند شد. به‌منظور سهولت کار، ورودی-های مجموعه در ماتریس $X = [x_1, x_2, \dots, x_n]$ و خروجی‌ها نیز در ماتریس $Y = [y_1, y_2, \dots, y_n]$ تجمیع می‌گردند. وظیفه رگرسیون، ایجاد یک ورودی جدید x^* به‌منظور دستیابی به توزیع پیش‌بینی شده برای مقادیر متناظر داده‌های مشاهداتی y^* و بر مبنای مجموعه داده S می‌باشد فرآیند گاوسی مجموعه‌ای از

متغیرهای تصادفی است که تعداد محدودی از آن‌ها با توزیع‌های گاوسی ادغام شده‌اند. فرآیند گاوسی تعمیمی از توزیع گاوسی می‌باشد. توزیع گاوسی در واقع توزیع بین متغیرهای تصادفی بوده در حالی که فرآیند گاوسی بیانگر توزیع بین توابع می‌باشد. فرآیند گاوسی $f(x)$ توسط توابع میانگین $m(x)$ و کوواریانس به‌شکل زیر تعریف می‌شود:

$$m(x) = E(f(x)) \quad (۸)$$

$$k(x, x') = E((f(x) - m(x))(f(x') - m(x')))) \quad (۹)$$

که در روابط فوق، $k(x, x')$ تابع کوواریانس (یا کرنل) بوده که در نقاط x و x' محاسبه می‌شود. فرایند گاوسی $f(x)$ می‌تواند به‌صورت زیر بیان گردد:

$$f(x) \sim GP(m(x), k(x, x')) \quad (۱۰)$$

که معمولاً جهت ساده‌سازی، مقدار تابع میانگین برابر با صفر در نظر گرفته می‌شود. در فرآیند گاوسی، رابطه بین بردار ورودی و هدف به فرم زیر می‌باشد:

$$y_i = f(x_i) + \varepsilon \quad (۱۱)$$

که در آن $f(x)$ بیانگر تابع رگرسیون دلخواه و ε نیز مقدار نویز توزیع گاوسی با میانگین صفر و واریانس σ^2 می‌باشد، یعنی علاوه بر این، چنین فرض می‌گردد که $[f(x_1), f(x_2), \dots, f(x_n)]^T$ رفتاری بر مبنای فرآیند گاوسی داشته به‌نحوی که $p(f|X) = N(0, K)$ که در آن K ماتریس کوواریانس با درایه‌های $k_{i,j} = k(x_i, x_j)$ می‌باشد.

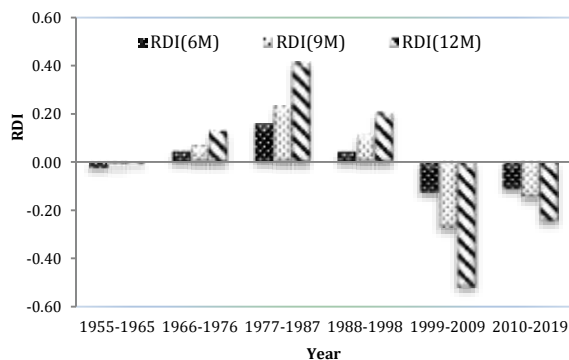
$$K(X, X) = \begin{pmatrix} k(x_1, x_1) & k(x_1, x_2) & \dots & k(x_1, x_n) \\ k(x_2, x_1) & k(x_2, x_2) & \dots & k(x_2, x_n) \\ \vdots & \vdots & \ddots & \vdots \\ k(x_n, x_1) & k(x_n, x_2) & \dots & k(x_n, x_n) \end{pmatrix} \quad (۱۲)$$

کوواریانس بین مقادیر توابع نهان $f(x_i)$ و $f(x_j)$ است. رگرسیون فرآیند گاوسی به‌منظور محاسبه توزیع پیش‌بینی شده برای مقادیر تابع f^* در نقاط تست $X^* = [x_1^*, x_2^*, \dots, x_m^*]$ به‌کار می‌رود. طراحی روش‌های رگرسیون مبتنی بر فرآیند گاوسی شامل استفاده از مفهوم تابع کرنل است. در واقع، با یک تبدیل غیرخطی از فضای ورودی به فضای خصیصه با ابعاد بیشتر (حتی نامتناهی) می‌توان مسائل را به‌صورت خطی تفکیک‌پذیر ساخت. با تبدیل نمونه‌ها از فضای ورودی به فضای ویژگی، تفکیک‌کننده غیرخطی به حالت خطی تبدیل خواهد شد. از مهم‌ترین توابع کرنل

جدول ۲- شاخص RDI محاسبه شده در سه مقیاس زمانی

ایستگاه	RDI مختلف		
	۱۲ (m)	۹ (m)	۶ (m)
اهر	-۵/۳۸۸	-۳/۱۰۵	-۲/۲۴۱
جلفا	-۲/۰۴۱	-۲/۴۷۵	۱/۹۶۳
تبریز	-۱/۰۲۳	-۱/۹۹	-۱/۲۴۱
مراغه	-۱/۳۹۸	-۱/۱۵۹	-۱/۴۹۲

شاخص RDI در سه مقیاس ۶ ماهه، ۹ ماهه و یک ساله برآورد گردید تا تغییرات ماهانه خشکسالی در طول دوره آماری مورد مطالعه را بهتر نشان دهد. پایش خشکسالی در دوره‌های ۱۰ ساله بین مقیاس‌های مختلف محاسباتی انجام گرفت. به منظور پرهیز از پُر حجم شدن مقاله در این بخش تنها نمودار ایستگاه مراغه ارائه شده است. نتایج محاسبات RDI (6M) و RDI (9M) نزدیک به هم بوده و اختلاف خیلی کمی باهم دارند در حالی که RDI (12M) نتایج متفاوتی به دست آورده است. طبق پژوهش احمدی و همکاران (۱۴۰۰) برای پایش خشکسالی حوضه درورد-بروجرد از شاخص RDI در مقیاس سالانه، ماهانه و منطقه مورد بررسی قرار دادند و نشان دادند مقیاس سالانه خشکسالی را بهتر نمایان می‌کند. بنابراین شاخص RDI در مقیاس سالانه برای ایستگاه مراغه ترسیم و در شکل (۲) ارائه شده است. با توجه به روند خشکسالی نشان داده شده در شکل (۲) از سال ۱۹۹۸ تا ۲۰۱۹ خشکسالی‌های شدید در حال وقوع می‌باشد. در سایر ایستگاه‌های مورد مطالعه نیز نتیجه به همین شکل بوده است. به طوری که در پایش ده سال اخیر از سال ۲۰۱۰ تا ۲۰۱۹ خشکسالی‌های پی‌پی با شدت‌های خفیف، متوسط و گاهی شدید به وقوع پیوسته‌اند. شکل (۳) شاخص RDI سالانه در ایستگاه مراغه را نشان می‌دهد. مطابق این شکل و جدول (۲) می‌توان چنین استنباط کرد که در سال‌های ۱۹۵۷ تا ۱۹۶۴، ۱۹۹۸ تا ۲۰۰۳، ۲۰۰۸ تا ۲۰۱۰ و ۲۰۱۶ تا ۲۰۱۸ خشکسالی خفیف تا خیلی شدید اتفاق افتاده است.



شکل ۲- نتایج محاسبه RDI در مقیاس‌های زمانی ۶، ۹ و ۱۲ ماهه برای ایستگاه مراغه

می‌توان به کرنل چند جمله‌ای ساده، چند جمله‌ای نرمال شده، تابع شعاع محور و تابع کرنل پیرسون اشاره کرد (روشنگر و همکاران، ۱۳۹۹). لازم به توضیح است که برای مدل‌سازی با روش‌های رگرسیون GPR و الگوریتم Kstar از نرم‌افزار Weka استفاده شده است.

۲-۶- ارزیابی مدل‌ها

در این تحقیق، برای ارزیابی مدل‌های موردنظر از معیارهای ضریب همبستگی (CC)، نش-سانکلیف (NS)، ریشه دوم میانگین مربعات خطا (RMSE) و میانگین خطای مطلق (MAE) استفاده می‌شود:

$$CC = \frac{\text{cov}(O_i, P_i)}{\sqrt{\text{Var}(O_i) \times \text{Var}(P_i)}} \quad (13)$$

$$NS = 1 - \frac{\sum_{i=1}^n (O_i - P_i)^2}{\sum_{i=1}^n (O_i - \bar{O})^2} \quad (14)$$

$$MAE = \frac{\sum_{i=1}^N |O_i - P_i|}{N} \quad (15)$$

$$RMSE = \sqrt{\frac{\sum (O_i - P_i)^2}{N}} \quad (16)$$

در روابط فوق، O_i و P_i به ترتیب مقادیر RDI محاسبه شده با داده‌های مشاهده‌ای، RDI پیش‌بینی شده با مدل‌های مذکور و میانگین RDI محاسبه شده با داده‌های مشاهده‌ای می‌باشند. CC ضریب همبستگی بین داده‌های محاسباتی و پیش‌بینی شده و N تعداد داده‌ها می‌باشد. مدلی به‌عنوان مناسب‌ترین گزینه انتخاب می‌گردد که بیشترین مقدار CC و کمترین مقدار RMSE، MAE و NS را به خود اختصاص دهد (احمدی و مداح، ۱۴۰۰).

۳- نتایج و بحث

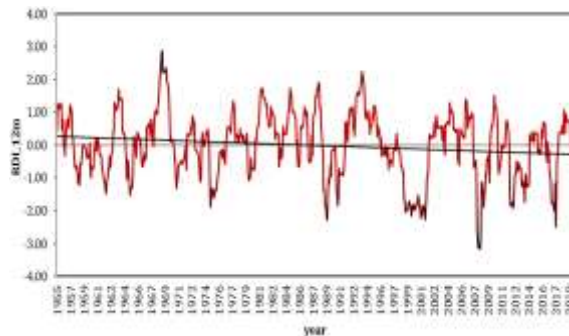
۳-۱- محاسبه شاخص RDI

در این مطالعه شاخص خشکسالی RDI برای چهار ایستگاه سینوپتیک منتخب شرق حوضه دریاچه ارومیه با استفاده از داده‌های هواشناسی بارش و تبخیر-تعرق پتانسیل محاسبه و مورد مقایسه قرار گرفت. نتایج حاصل در جدول (۲) ارائه گردیده است. شاخص RDI در مقاطع زمانی ۶، ۹ و ۱۲ ماهه محاسبه شد، نتایج به‌دست آمده از کفایت بالایی در نشان دادن خشکسالی برخوردار می‌باشند. با توجه به نتایج حاصل از بین مقیاس‌های زمانی، شاخص RDI یک‌ساله (۱۲ ماهه) بهتر می‌تواند خشکسالی رخ داده را بازگو کند.

پراکندگی به دست آمده در سه مقیاس زمانی مختلف ماهانه تغییرات را نشان می‌دهند و با توجه به آن‌ها می‌توان مشاهده کرد که الگوریتم‌های به کار گرفته در مقیاس ۱۲ ماهه به خوبی توانسته‌اند شاخص RDI را مدل کنند. همچنین از بین سه مقیاس زمانی مورد استفاده مقیاس ۱۲ ماهه به خوبی و با دقت بالا خشکسالی رخ داده شده را نمایان می‌کند. بنابراین طبق نتیجه پژوهش Mehdizadeh و همکاران (۲۰۲۰) مشابه با تحقیق حاضر می‌توان ادعا کرد پایش سالانه خشکسالی نتایج دقیق تری نسبت به پایش ماهانه دارند.

برای بررسی عملکرد مدل‌های منفرد به کار گرفته شده در تحقیق حاضر، با کمک نرم‌افزار متلب معیارهای ارزیابی برای چهار ایستگاه مورد مطالعه در سه مقیاس زمانی ذکر شده مورد محاسبه و بررسی قرار گرفتند. در جدول (۳) اختلاف بسیار کمی بین معیارهای ارزیابی دو الگوریتم GPR و Kstar دیده می‌شود که حاکی از عملکرد و دقت مدل‌سازی بالای هر دو مدل می‌باشد. ولی با این وجود مقایسه بین آن‌ها نشان می‌دهد که الگوریتم GPR بهتر توانسته شاخص RDI را مدل کند و پراکندگی و خطای موجود را از بین ببرد. به طوری که با توجه به جدول (۳)، برای بررسی عملکرد الگوریتم‌ها می‌توان از نتیجه محاسبه ضریب همبستگی استفاده کرد. ضریب همبستگی به ترتیب در مقیاس زمانی یک‌ساله برای الگوریتم GPR مقدار ۰/۹۲ و برای الگوریتم Kstar بین مقادیر ۰/۸۲ تا ۰/۹۲ بوده است. همچنین در مقیاس ۶ و ۹ ماهه ضریب همبستگی برای الگوریتم GPR بین ۰/۸۲ تا ۰/۸۶ و در الگوریتم Kstar بین ۰/۸۲ تا ۰/۸۵ می‌باشد.

در بررسی دقت الگوریتم‌های مورد استفاده نیز از نتیجه معیارهای ارزیابی جذر میانگین مربعات خطا، میانگین مطلق خطا و نش-ساتکلیف^{۱۱} می‌توان بهره‌مند شد. با توجه به جدول (۳)، مقدار عددی RMSE در مقیاس زمانی یک‌ساله با الگوریتم GPR در بازه ۰/۳۱ تا ۰/۳۹ و با الگوریتم Kstar در بازه ۰/۳۲ تا ۰/۵۱ قرار دارد. همچنین در مقیاس‌های ۶ و ۹ ماهه نیز به ترتیب با الگوریتم GPR، ۰/۵۱ تا ۰/۵۵ و با الگوریتم Kstar، ۰/۵۲ تا ۰/۵۷ شاخص خشکسالی RDI را مدل‌سازی کردند. با توجه به نتایج ذکر شده می‌توان از الگوریتم GPR با دقت بالا در مدل‌سازی استفاده کرد. همین‌طور می‌توان بیان کرد که مقیاس زمانی یک‌ساله در مقایسه با مقیاس‌های ماهانه بهتر و واقعی‌تر خشکسالی را نشان می‌دهد.



شکل ۳- پایش خشکسالی ایستگاه مراغه در طول دوره آماری مورد مطالعه با RDI سالانه

۳-۲- مدل‌سازی شاخص RDI با مدل‌های منفرد Kstar و GPR

در این مطالعه به منظور پیش‌بینی خشکسالی ماهانه ایستگاه‌های مورد مطالعه، ۸۰ درصد داده‌ها (۵۳۶ ماه) برای آموزش و ۲۰ درصد مابقی برای آزمون (۲۳۰ ماه) در نظر گرفته شد. برای توسعه الگوریتم‌های Kstar و GPR از نرم‌افزار WEKA استفاده شد و پارامترهای مورد استفاده آن از روش آزمون و خطا تعیین گردید. توابع کرنل با توجه به ساختار ریاضی، نیازمند تعریف و انتخاب عواملی هستند که برای انتخاب آن‌ها از روش سعی و خطا استفاده می‌شود. هنگام استفاده از تابع مختلف کرنل مقدار پارامتر نویز^۹ فرایند گاوسی بین ۰/۱ تا ۱ در نظر گرفته شد و ملاحظه گردید که با افزایش مقدار نویز نتایج تغییر نمی‌کنند و بهترین نتیجه در این محدوده حاصل می‌شود. همچنین در استفاده از تابع کرنل پایه شعاعی مقدار گاما بین ۰/۱ و ۳ بررسی و مشاهده گردید که بهترین نتیجه بین مقادیر ۰/۱ و ۲ به دست آمد. جودی و ستاری (۱۳۹۴) نیز نتایج مشابهی را در خصوص افزایش پارامترهای مؤثر در مدل GPR گزارش نمودند.

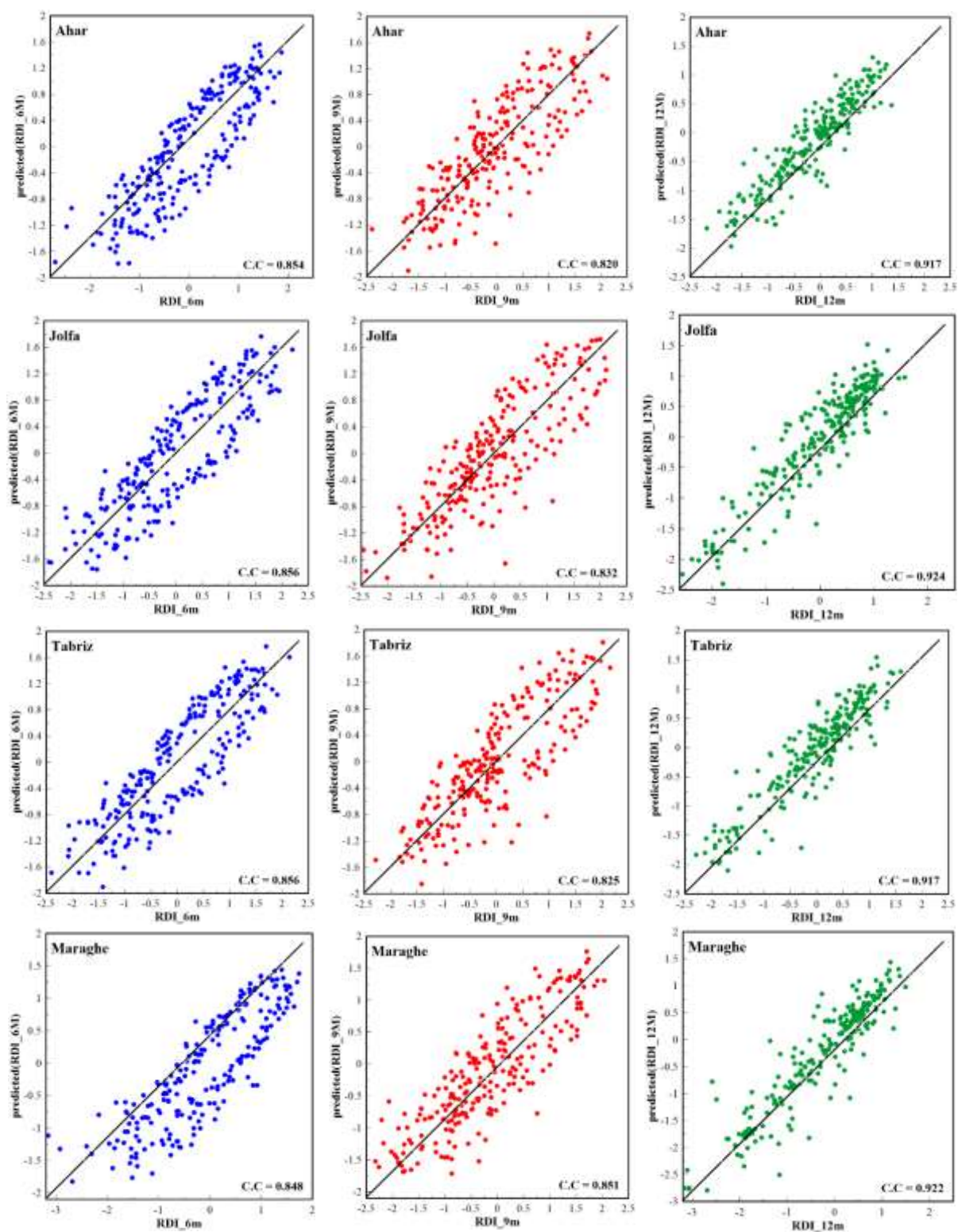
در الگوریتم Kstar پارامتر ترکیب جهانی^{۱۰} که بین ۱۰ تا ۱۰۰ متغیر است در این پژوهش برابر با ۱۰ به دست آمد. لازم به ذکر است که با افزایش پارامتر ترکیب جهانی عملکرد مدل در برآورد شاخص خشکسالی کاهش می‌یابد.

احمدی و مداح (۱۴۰۰) نیز نتایج مشابهی در این خصوص ارائه نموده‌اند. با توجه به این‌که در داده‌ها خلأ آماری وجود نداشت از این‌رو حالت گمشده^{۱۱} در الگوریتم را به صورت پیش فرض انتخاب و تغییری در آن صورت نگرفت.

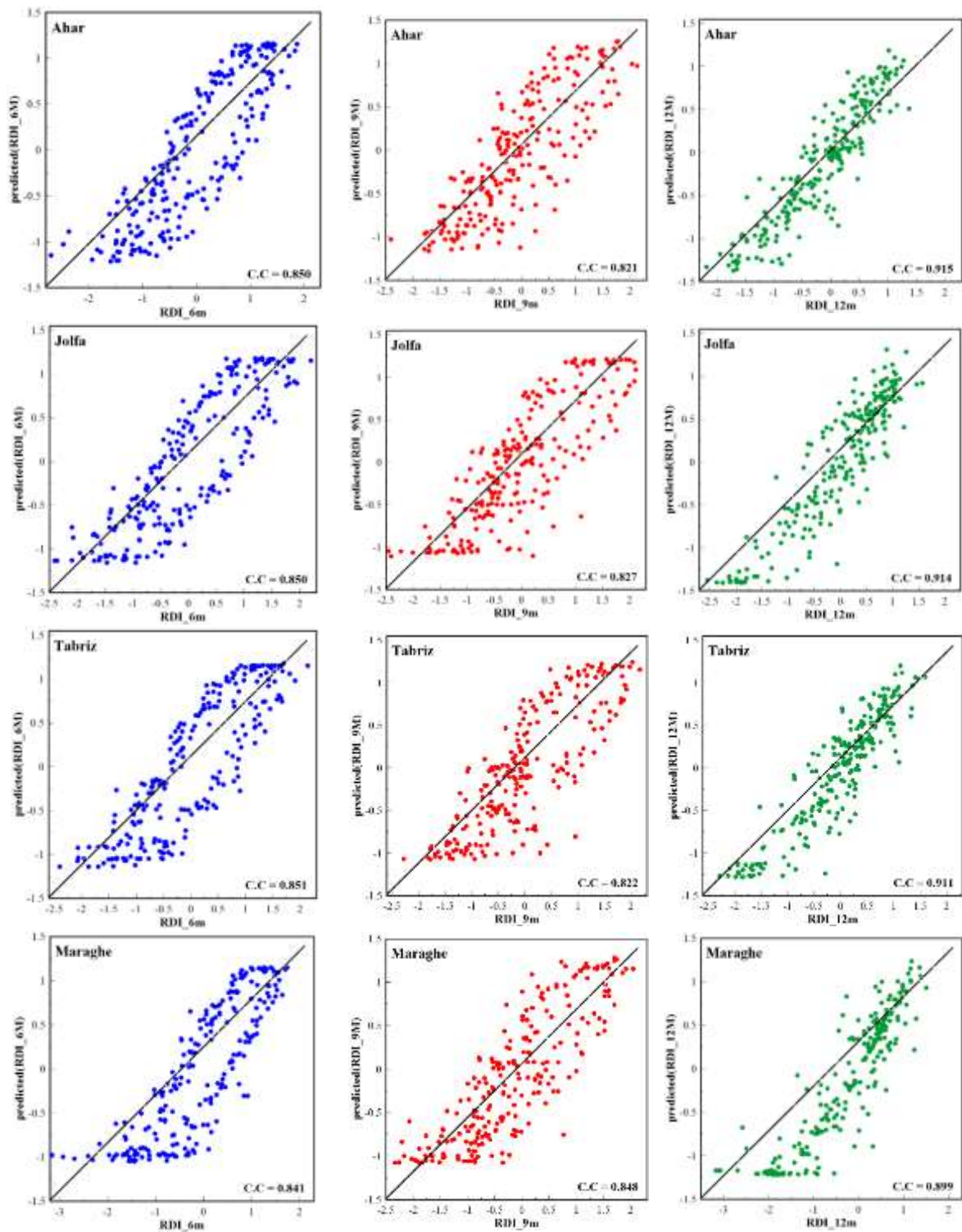
نتایج حاصل از مدل‌سازی شاخص RDI با الگوریتم‌های ذکر شده در شکل‌های (۴) و (۵) آورده شده است. نمودارهای

11. Missing Mode
12. Nash-Sutcliffe efficiency

9. Noise
10. Global Blend



شکل ۴- مدل سازی شاخص RDI با مدل GPR در ایستگاه های شرق دریاچه ارومیه



شکل ۵- مدل سازی شاخص RDI با مدل Kstar در ایستگاه های شرق دریاچه ارومیه

جدول ۳- معیارهای ارزیابی عملکرد مدل‌های مورد استفاده در مدل‌سازی شاخص RDI

معیارهای ارزیابی				مقیاس زمانی	مدل مورد استفاده	ایستگاه
CC	NS	MAE	RMSE			
۰/۸۵	۰/۷۳	۰/۴۴	۰/۵۱	۶ ماهه	GPR	اهر
۰/۸۲	۰/۶۷	۰/۴۲	۰/۵۳	۹ ماهه		
۰/۹۲	۰/۸۴	۰/۲۳	۰/۳۱	۱۲ ماهه		
۰/۸۶	۰/۷۳	۰/۴۵	۰/۵۲	۶ ماهه	GPR	تبریز
۰/۸۲	۰/۶۸	۰/۴۴	۰/۵۴	۹ ماهه		
۰/۹۲	۰/۸۴	۰/۲۵	۰/۳۴	۱۲ ماهه		
۰/۸۶	۰/۷۳	۰/۴۶	۰/۵۳	۶ ماهه	GPR	جلفا
۰/۸۳	۰/۶۹	۰/۴۴	۰/۵۵	۹ ماهه		
۰/۹۲	۰/۸۵	۰/۲۶	۰/۳۵	۱۲ ماهه		
۰/۸۵	۰/۷۲	۰/۴۶	۰/۵۴	۶ ماهه	GPR	مراغه
۰/۸۵	۰/۷۲	۰/۴۴	۰/۵۳	۹ ماهه		
۰/۹۲	۰/۸۵	۰/۲۶	۰/۳۹	۱۲ ماهه		
۰/۸۵	۰/۷۲	۰/۴۴	۰/۵۲	۶ ماهه	Kstar	اهر
۰/۸۲	۰/۶۷	۰/۴۲	۰/۵۳	۹ ماهه		
۰/۹۲	۰/۸۳	۰/۲۴	۰/۳۲	۱۲ ماهه		
۰/۸۵	۰/۷۲	۰/۴۶	۰/۵۳	۶ ماهه	Kstar	تبریز
۰/۸۲	۰/۶۷	۰/۴۵	۰/۵۶	۹ ماهه		
۰/۹۱	۰/۸۱	۰/۲۸	۰/۳۷	۱۲ ماهه		
۰/۸۵	۰/۷۱	۰/۴۸	۰/۵۶	۶ ماهه	Kstar	جلفا
۰/۸۳	۰/۶۷	۰/۴۶	۰/۵۷	۹ ماهه		
۰/۹۱	۰/۸۲	۰/۲۹	۰/۳۸	۱۲ ماهه		
۰/۸۴	۰/۶۹	۰/۴۸	۰/۵۷	۶ ماهه	Kstar	مراغه
۰/۸۵	۰/۷۰	۰/۴۶	۰/۵۶	۹ ماهه		
۰/۹۰	۰/۷۵	۰/۳۴	۰/۵۱	۱۲ ماهه		

۴- نتیجه‌گیری

این مطالعه با هدف مدل‌سازی خشکسالی RDI در سه مقیاس زمانی شامل RDI-6m، RDI-9m و RDI-12m برای چهار ایستگاه شرق حوضه دریاچه ارومیه انجام شد. برای دستیابی به این هدف، از دو الگوریتم، Kstar و GPR برای مدل‌سازی شاخص RDI خشکسالی استفاده شد و دقت آن‌ها مورد ارزیابی قرار گرفت. بررسی نتایج حاصل از محاسبه شاخص RDI با کلاس‌بندی ارائه شده توسط Tsakiris (۲۰۰۷) حاکی از آن بود که در طی سال‌های مورد مطالعه بخش شرق دریاچه ارومیه از نظر خشکسالی وضعیت "نسبتاً خشک" و "به شدت خشک" را دارد.

در مدل‌سازی شاخص RDI از ۸۰ درصد داده‌ها برای آموزش و ۲۰ درصد مابقی داده‌ها برای آزمون مدل مورد استفاده قرار گرفت. نتایج حاصل نشان داد که مدل‌های منفرد Kstar و GPR در برآورد شاخص RDI ماهانه از کارایی لازم و عملکرد بالایی برخوردار بوده است. به طوری که در نمودارهای پراکندگی به دست آمده مشاهده گردید که الگوریتم‌های ذکر شده به خوبی توانسته‌اند پراکندگی داده‌ها را از بین برده و نتایج قابل قبولی ارائه دهند. سپس برای مقایسه و انتخاب الگوریتم قوی‌تر، از معیارهای ارزیابی ضریب همبستگی، ضریب نش-ساتکلیف، جذر میانگین مربعات خطا و میانگین مطلق خطا استفاده شد. نتایج حاصل از مقادیر

عددی معیارهای محاسبه‌ای نشان داد در مقیاس زمانی یک‌ساله (RDI-12m) ضریب همبستگی عدد بیشتر و معیارهای خطا عدد کمتری گزارش کردند. همچنین از بین دو الگوریتم مورد استفاده، الگوریتم GPR با اختلاف بسیار کمی از Kstar توانسته است شاخص RDI را مدل‌سازی کند.

۵- مراجع

احمدی ف، "ارزیابی عملکرد روش‌های ماشین‌بردار پشتیبان و سیستم استنتاج عصبی فازی تطبیقی در پیش‌بینی جریان ماهانه رودخانه (مطالعه موردی: رودخانه‌های نازلو و سزار)،" تحقیقات آب و خاک ایران، ۱۳۹۹، ۵۱ (۳)، ۶۸۶-۶۷۳.

احمدی ف، مداح م، "توسعه روش هیبریدی موجک-الگوریتم Kstar برای پیش‌بینی بارش‌های ماهانه (مطالعه موردی: ایستگاه سینوپتیک اهواز)،" تحقیقات آب و خاک ایران، ۱۴۰۰، ۵۲ (۲)، ۴۲۰-۴۰۹.

احمدی م، پایمزد ش، رحیمی م، "مقایسه پایش خشکسالی هیدرولوژی و هواشناسی با استفاده از شاخص‌های RDI و EDI"، محیط زیست و مهندسی آب، ۱۴۰۰، ۷ (۴)، ۶۹۶-۶۸۳.

- Mehdizadeh S, Ahmadi F, Danandeh Mehr A, Jafar Sadegh Safari M, "Drought modeling using classic time series and hybrid wavelet-gene expression programming models", *Journal of Hydrology*, 2020. doi:https://doi.org/10.1016/j.jhydrol.2020.125017
- Mishra A, Singh VP, "A review of drought concepts", *Journal of Hydrology*, 2010, 391, 202-216.
- Palmer WC, "Meteorological Drought US Department of Commerce", Weather Bureau, Washington, DC, 1965.
- Tsakiris G, Pangalou D, Vangelis H, "Regional drought assessment based on the reconnaissance drought index (RDI)", *Water Resource Management*, 2007, 21, 821-833.
- Vicente-Serrano SM, Beguería S, López-Moreno JI, "A multi scalar drought index sensitive to global warming: the standardized precipitation evapotranspiration index", *Journal of Climatology*, 2010, 23, 1696-1718.
- Wilhite DA, Glantz MH, "Understanding: the drought phenomenon", *The Role of Definitions Water*, 1985, 10, 111-120.
- شکوهی ع، "مقایسه شاخص‌های RDI و SPI برای تحلیل خشکسالی در مقیاس ایستگاهی با تکیه بر خشکسالی کشاورزی (مطالعه موردی: قزوین و تاکستان)"، فصلنامه علمی- پژوهشی مهندسی آبیاری و آب، ۱۳۹۱، ۹ (۳)، ۱۲۲-۱۱۱.
- معینی فر س، اسدی م، ا، ملکی نژاد ح، طالبی ع، "تعیین توزیع آماری مناسب برای محاسبه شاخص RDI در مناطق خشک (مطالعه موردی: ایران مرکزی)"، نشریه علمی خشک بوم، ۱۴۰۰، ۱۱۱ (۱)، ۱۲۱-۱۰۵.
- رضازاده جودی ع، ستاری م ت، "تخمین عمق چاله آبستگي پایه پل در سازه‌های آبی با روش رگرسیون فرایند گاوسی"، مجله تحقیقات کاربردی مهندسی سازه‌های آبیاری و زهکشی، ۱۳۹۴، ۶۵ (۱۶)، ۳۶-۱۹.
- میرعباسی نجف‌آبادی ر، احمدی ف، عاشوری م، ناظری تهرودی م، "تحلیل خشکسالی‌های شمال شرق ایران با استفاده از شاخص کمبود توأم (JDI)"، اکوهیدرولوژی، ۱۳۹۶، ۴ (۲)، ۵۸۵-۵۷۳.
- فاطمی م، رحیمیان م، ح، اکرامی م، برخوردار ج، "تحلیل فضایی نمایه خشکسالی اجمالی (RDI) در ایران مرکزی (مطالعه موردی: استان یزد)"، نشریه علمی- پژوهشی مهندسی آبیاری و آب ایران، ۱۳۹۸، ۳۶ (۹)، ۱۶۷-۱۶۰.
- Beyaztas U, Yaseen ZM, "Drought interval simulation using functional data analysis", *Journal of Hydrology*, 2019, 579, 124141.
- Ekmekcioğlu Ö, Başakın EE, Özger M, "Tree-based nonlinear ensemble technique to predict energy dissipation in stepped spillways", *European Journal of Environmental and Civil Engineering*, 2020, 1-19.
- Granata F, Di Nunno F, Gargano R, de Marinis G, "Equivalent discharge coefficient of side weirs in circular channel- a lazy machine learning approach", 2019, *Water*, 11 (11), 2406.
- Khadr M, "Forecasting of meteorological drought using Hidden Markov Model (case study: The upper Blue Nile river basin, Ethiopia)", *Ain Shams Engineering Journal*, 2016, 7 (1), 47-56.
- Li Q, He P, He Y, Han X, Zeng T, Lu G, Wang H, "Investigation to the relation between meteorological drought and hydrological drought in the upper Shaying River Basin using wavelet analysis", *Atmospheric Research*, 2020, 234,
- Madadgar S, Moradkhani H, "Spatio-temporal drought forecasting within Bayesian networks", *Journal of Hydrology*, 2014, 512, 134-146.
- Madrigal J, Solera A, Suárez-Almiñana S, Paredes-Arquiola J, Andreu J, Sánchez-Quispe ST, "Skill assessment of a seasonal forecast model to predict drought events for water resource systems", *Journal of Hydrology*, 2018, 564, 574-587.
- Mckee TB, Doesken NJ, Kleist J, "The relationship of drought frequency and duration to time scales", *AMS 8th Conference, Climatology*, 1993, 179-184.

EXTENDED ABSTRACT

Investigating the Performance of Kstar and GPR Algorithms in Modeling RDI Meteorological Drought Index (Case Study: East of Urmia Lake Basin)

Marzie Sadeghian Aghkandi^a, Hossein Rezaie^{a,*}, Keivan Khalili^a, Farshad Ahmadi^b

^a Department of Water Science and Engineering, Urmia University, Iran

^b Department of Hydrology and Water Resources, Shahid Chamran University of Ahvaz, Iran

Received: 13 February 2023; Review: 13 April 2023; Accepted: 17 May 2023

Keywords:

Kstar algorithm, RDI drought monitoring, Gaussian process regression, east of Lake Urmia.

1. Introduction

Drought is a severe hydrological event that can cause serious problems in human life. In this context, it can have adverse effects on water supply and quality, public health, agricultural productivity, land degradation, desertification, famine, etc. (Madadgar and Moradkhani, 2014; Li et al., 2020). In a general classification, drought events are classified into four different categories, meteorological, hydrological, agricultural and socio-economic droughts (Wilhite and Glantz, 1985; Khadr, 2016). Some of the well-known and common meteorological drought indices in drought monitoring include Palmer Drought Severity Index (PDSI), Drought Identification Index (RDI), Standardized Precipitation Index (SPI) and Standard Precipitation Evapotranspiration Index.

Considering the significant role of evaporation and transpiration in the water balance, it is necessary to consider its effect when studying drought in a particular region. Based on this, RDI, which includes both precipitation and evaporation and transpiration, can be considered as a reliable indicator for drought monitoring (Moeinifar et al., 1400). According to importance of drought as a natural phenomenon in hydrological and meteorological studies, its monitoring and forecasting with a suitable approach can be important. The main role of Drought prediction in risk management, reducing the effects of drought on existing water resources and their optimal use, the possibility of rational decision-making by decision makers to minimize the damages caused by drought, as well as planning and managing resource projects. It has water. (Khadr et al., 2016; Madrigal et al., 2018; Beyaztas and Yaseen, 2019). Among the models worked till date, single Kstar and GPR models are the newest models for drought prediction

2. Methodology

2.1. Data and the study area

It is located in the east of Urmia lake basin with an area of 1084.38 square kilometers between the north latitude of minimum 36 and maximum 38 degrees and east longitude of minimum 46 and maximum 48 degrees in the eastern-most part of the catchment area of Lake Urmia. In this range, the incoming flows are; Surface currents, rainfall and snow and permanent channels. The average annual rainfall in the plains of this area is calculated to be 330 mm. The meteorological information used in this research is the monthly rainfall and potential evapotranspiration data of Ahar, Tabriz, Jolfa and Maragheh synoptic stations in the years 1955 to 2019 to estimate the RDI index.

* Corresponding Author

E-mail addresses: m.sadeghian@urmia.ac.ir (Marzie Sadeghian Aghkandi), h.rezaie@urmia.ac.ir (Hossein Rezaie), k.khalili@urmia.ac.ir (Keivan Khalili), f.ahmadi@scu.ac.ir (Farshad Ahmadi).

2.2. RDI drought recognition index

The RDI index is more popular than other indices due to the combined consideration of precipitation and potential evaporation and transpiration, and in many studies, it is the basis for comparing the severity of drought and estimating its potential damages at the national level. The RDI index with the same structure as SPI has special capabilities for evaluating agricultural and hydrological drought due to the use of potential evaporation and transpiration in its structure (Shkohi, 2013). In the studied area, the performance comparison of Kstar and GPR models on the RDI index has not been done so far. Therefore, according to the capabilities of these two models, in this research, meteorological drought monitoring in the eastern basin of Lake Urmia in a period of 65 years, and the analysis of the results of these two models in the RDI index were considered.

2.3. Kstar algorithm

The Kstar algorithm can be defined as a method for cluster analysis whose main goal is to divide n observations into K clusters so that each observation belongs to the cluster with the closest average (Ekmekcioğlu et al, 2020). Kstar algorithm can be described as an example-based learning method that uses entropy theory as a distance measure. This method maximizes the possibility of obtaining information from existing data by providing a consistent approach for managing real valuable features, symbolic features and missing values. In the Kstar algorithm, the distance from one sample to another can be described as the complexity of converting one sample to another (Granata et al, 2019). As mentioned, Kstar regression uses an entropy distance function. Entropic distance is used to obtain samples that are most similar to each other from the data set. Consider.

2.4. Gaussian Process Regression (GPR)

In the theory of statistics and probability, a Gaussian process is a statistical model in which observations occur in a continuous range. In fact, the Gaussian process is a random process that is known as a sequence of random variables, and the random variables are usually calculated as; For example, time is ordered and each finite subset of these variables has a multivariate Gaussian (normal) distribution. In a Gaussian process, every point of the input space is a random variable with a normal distribution. In addition, each finite set of these random variables has a multivariate Gaussian distribution. Gaussian process distribution is the joint distribution of all these random variables (number and infinite). From the point of view of a machine learning algorithm, a Gaussian process measures the similarity between points (the same kernel function) to predict new points from the training data. Gaussian process is a set of random variables, a limited number of which are integrated with Gaussian distributions. Gaussian process is completely determined by its mean function $m(x)$ and its covariance function $k(x,x)$. This process is a natural generalization of the Gaussian distribution, whose mean and covariance are a vector and a matrix, respectively. Gaussian process regression models are based on the assumption that the set observations should carry information about each other. One of the important features of the Gaussian process is the presence of various covariance functions in it, which allows the researcher to make an appropriate selection among them. These models can specify distributions between functions with one or more input variables. When such a function defines the mean response in a regression model with Gaussian errors, matrix calculations can be used for inference. - got stuck; This problem is possible for data sets with more than a thousand samples

3. Results and discussion

3.1. Results

In this study, RDI drought index was calculated and compared for four selected synoptic stations in the east of Urmia lake basin in six, nine and 12 months using meteorological data of precipitation and potential evapotranspiration. The obtained results have high adequacy in showing drought. According to the results obtained from different time scales, the one-year RDI index (12 months) can better describe the drought that occurred. Also, drought monitoring was done in 10-year periods between different calculation scales. The calculation results of RDI (6M) and RDI (9M) are close to each other and have very little difference, while RDI (12M) has obtained different results. According to Ahmadi et al.'s research (1400), they examined the RDI index in the annual, monthly and regional scales to monitor the drought in Durward-Broujerd basin and showed that the annual scale shows the drought better. According to the obtained results, it can be concluded that in the years 1957 to 1964, 1998 to 2003, 2008 to 2010 and 2016 to 2018, mild to very severe drought occurred. In this study, WEKA software was used to develop Kstar and GPR algorithms, and its parameters were determined

by trial and error. In order to predict the monthly drought of the studied stations, 80% of the data (536 months) were considered for training and the remaining 20% for the test (230 months).

In checking the accuracy of the used algorithms, the results of the root mean square error, absolute mean error and Nash-Sutcliff evaluation criteria can be used. The numerical value of RMSE in a one-year time scale with the GPR algorithm is in the range of 0.31 to 0.39 and with the Kstar algorithm in the range of 0.32 to 0.51. Also, in the scales of 6 and 9 months, they modeled RDI drought index from 0.51 to 0.55 with GPR algorithm and from 0.52 to 0.57 with Kstar algorithm. According to the mentioned results, the GPR algorithm can be used with high accuracy in modeling. It can also be said that the one-year time scale shows the drought better and more realistically compared to the monthly scales.

4. Conclusions

This study was conducted with the aim of modeling RDI drought in three time scales including RDI-6m, RDI-9m and RDI-12m for four stations east of Urmia lake basin. To achieve this goal, two algorithms, Kstar and GPR, were used to model the drought RDI index and their accuracy was evaluated. Examining the results of calculating the RDI index with the classification presented by Tsakiris et al., (2007) indicated that during the studied years, the eastern part of Lake Urmia was "relatively dry" and "extremely dry" in terms of drought. Dry"

The results showed that the individual Kstar and GPR models have the necessary efficiency and high performance in estimating the monthly RDI index. As shown in the obtained scatter diagrams, it was observed that the mentioned algorithms were able to eliminate the scatter of the data and provide acceptable results. Then, to compare and select a stronger algorithm, the evaluation criteria of correlation coefficient, Nash-Sutcliffe coefficient, root mean square error and absolute mean error were used. The results obtained from the numerical values of the calculation criteria showed that in the one-year time scale (RDI-12m), the correlation coefficient was higher and the error criteria were lower. Also, among the two used algorithms, the GPR algorithm has been able to model the RDI index with a very small difference from Kstar.

5. References

- Ahmadi F, Madah MA, "Development of hybrid method of wavelet-Kstar algorithm for forecasting monthly rainfall (case study: Ahvaz synoptic station)", *Iran Water and Soil Research*, 1400, 52 (2), 409-420 (In Persian).
- Beyaztas U, Yaseen ZM, "Drought interval simulation using functional data analysis", *Journal of Hydrology*, 2019, 579, 124141.
- Granata F, Di Nunno F, Gargano R, de Marinis G, "Equivalent discharge coefficient of side weirs in circular channel- a lazy machine learning approach", 2019, *Water*, 11 (11), 2406.
- Khadr M, "Forecasting of meteorological drought using Hidden Markov Model (Case Study: The upper Blue Nile river basin, Ethiopia)", *Ain Shams Eng Journa*, 2016, 7 (1), 47-56.
- Li Q, He P, He Y, Han X, Zeng T, Lu G, Wang H, "Investigation to the relation between meteorological drought and hydrological drought in the upper Shaying River Basin using wavelet analysis", 2020, *Atmos. Res.* 234, 104743.
- Madadgar S, Moradkhani H, "Spatio-temporal drought forecasting within Bayesian networks", *Journal of Hydrology*, 2014, 512, 134-146.
- Madrigal J, Solera A, Suárez-Almiñana S, Paredes-Arquiola J, Andreu J, Sánchez-Quispe ST, "Skillassessment of a seasonal forecast model to predict drought events for water resource systems", *Journal of Hydrology*, 2018, 564, 574-587.
- Mckee TB, Doesken NJ, Kleist J, "The relationship of drought frequency and duration to time scales", *AMS 8th Conference, Climatology*, 1993, 179-184.
- Moinifar S, Asadi MA, Melki-Nejad H, Talebi A, "Determining the appropriate statistical distribution for calculating the RDI index in arid regions (Case Study: central Iran)", *Arid Scientific Journal*, 1400, 11 (1), 105-121 (In persian).
- Mirabbasi Najafabadi R, Ahmadi F, Ashuri M, Nazeri Tehrodi M, "Analysis of droughts in North-Eastern Iran using Joint Deficit Index (JDI)", *Ecohydrology*, 2016, 4 (2), 573-585 (In persian).
- Palmer WC, "Meteorological Drought US Department of Commerce", *Weather Bureau*, Washington, DC, 1965.
- Shokohi A, "Comparison of RDI and SPI indices for station-scale drought analysis based on agricultural drought (case study: Qazvin and Takestan)", *Scientific Research Journal of Irrigation and Water Engineering*, 1391, 9 (3), 111-122.
- Tsakiris G, Pangalou D, Vangelis H, "Regional drought assessment based on the reconnaissance drought index (RDI)", *Water Resource, Management*, 2007, 21, 821-833.

Vicente-Serrano SM, Beguería S, López-Moreno JI, "A multi scalar drought index sensitive to global warming: the standardized precipitation evapotranspiration index", *Journal of Climatology*, 2010, 23, 1696-1718.
Wilhite DA, Glantz MH, "Understanding: the drought phenomenon", 1985, *the Role of Definitions Water Int.* 10, 111-120.