

ارزیابی خشکسالی به کمک شاخص بارش استاندارد و الگوریتم جنگل تصادفی

زینب ترابی^۱، رمضان واقعی^{۲*}، علیرضا قائمی^۳

^۱ دانشجوی کارشناسی ارشد مهندسی عمران- محیط زیست، دانشکده مهندسی عمران، دانشگاه صنعتی شاهرود

^۲ استادیار گروه مهندسی آب و محیط زیست، دانشکده مهندسی عمران، دانشگاه صنعتی شاهرود

^۳ دانشجوی دکتری مهندسی عمران گرایش مهندسی و مدیریت منابع آب، دانشگاه سیستان و بلوچستان

تاریخ دریافت: ۹۹/۰۶/۲۶، تاریخ پذیرش: ۹۹/۰۹/۲۷

چکیده

خشکسالی باعث تشدید بحران آب و ایجاد خسارت های جبران ناپذیر به جوامع می شود. در سال های اخیر استفاده از روش های یادگیری ماشین در ارزیابی خشکسالی مورد توجه محققین قرار گرفته است. هدف این تحقیق، ارزیابی خشکسالی در شهرهای زابل و زاهدان طی بازه زمانی (۲۰۲۰-۱۹۹۰) است که برای این منظور شاخص بارش استاندارد (SPI) در مقیاس فصلی و سالانه، تحلیل خودهمبستگی جزئی (PACF) و الگوریتم جنگل تصادفی (RF) استفاده شد. پس از محاسبه SPI، نتایج تحلیل PACF برای SPI، به عنوان ورودی های مدل در نظر گرفته شد. داده های آموزش و آزمایش با دو ورودی متفاوت مورد بررسی قرار گرفت. طبق نتایج تحقیق در مقیاس فصلی و سالانه SPI، طبقه خشکسالی تقریباً نرمال (N) بیشترین فراوانی وقوع را در هر دو ایستگاه دارد و بر اساس تحلیل PACF، مناطق مورد مطالعه طی بازه زمانی ۳۰ ساله دچار تغییر اقلیم شده اند. نتایج مدل توسط چند پارامتر آماری بررسی شده اند. شاخص توافق (IOA) در ایستگاه زابل برای داده های آموزش با در نظر گرفتن چهار تاخیر زمانی (۲، ۴، ۶ و ۱۲ ماهه) و سه تاخیر زمانی (۲، ۴ و ۶ ماهه) به عنوان ورودی به ترتیب برابر ۰/۹۶۴۸ و ۰/۹۲۵۶ بوده و برای داده های آزمایش به ترتیب ۰/۸۵۵۶ و ۰/۸۶۷۳ می باشد. IOA در ایستگاه زاهدان برای داده های آموزش با چهار تاخیر زمانی (۲، ۴، ۶ و ۸ ماهه) و سه تاخیر زمانی (۲، ۴ و ۶ ماهه) به ترتیب برابر ۰/۹۴۹۵ و ۰/۹۲۰۵ است و برای داده های آزمایش به ترتیب ۰/۷۴۰۸ و ۰/۶۳۰۳ می باشد. سایر پارامتر های آماری بررسی شده نیز نشان دهنده قابلیت مطلوب RF در تخمین SPI می باشند.

کلمات کلیدی: خشکسالی، شاخص بارش استاندارد، الگوریتم جنگل تصادفی، تحلیل خودهمبستگی جزئی

طبق گزارش ارزیابی پنجم شورای بین المللی تغییر اقلیم، از دهه ۱۹۵۰ کره زمین به صورت فزاینده ای در حال گرم شدن است و خطر خشکسالی بیش از ۴۵ درصد سرزمین های جهان را تهدید می کند (لی و همکاران، ۲۰۱۹). از زمان انقلاب صنعتی با سوزاندن سوخت های فسیلی میزان تولید گازهای گلخانه ای عمدتاً دی اکسید کربن افزایش یافته است. تجمع دی اکسید کربن در اتمسفر باعث به دام افتادن تابش های ساطع شده از سطح زمین شده و در نهایت با گرم شدن کره زمین، تغییرات آب و هوایی ایجاد می گردد. با توجه به محدودیت منابع آب، رشد جمعیت و افزایش تقاضای آب برای مصارف گوناگون، بحران تامین آب ایجاد می شود که خشکسالی باعث تشدید این بحران می گردد (زندلینز و همکاران، ۲۰۲۱). خشکسالی پدیده ای طبیعی است و هنگامی رخ می دهد که منطقه ای در درازمدت دچار کاهش بارندگی شود (چن و همکاران، ۲۰۰۹). خشکسالی منجر به ایجاد خسارت های جبران ناپذیر در زمینه پایداری اجتماع، اقتصاد و محیط زیست می شود (شارما و همکاران، ۲۰۰۹). شاخص های متنوعی برای ارزیابی خشکسالی توسعه داده شده اند که متغیرهای اقلیمی و هیدرولوژیکی در محاسبه شاخص های خشکسالی رایج تر می باشند (هایم، ۲۰۰۲). بارندگی، دمای هوا، رطوبت خاک، تبخیر- تعرق و جریان در بیشتر روش ها ملاک تعیین شاخص های خشکسالی می باشند (چوی و همکاران، ۲۰۱۳). در سال های اخیر بکارگیری روش های یادگیری ماشین در مورد شاخص های خشکسالی بسیار مورد توجه قرار گرفته است. از جمله روش های مربوط به پایش خشکسالی می توان به الگوریتم جنگل تصادفی اشاره نمود. پرمادیا و اونیشی (۲۰۱۸) شاخص بارش استاندارد^۱ (SPI) را به صورت ۳،۶،۹،۱۲ ماهه در شهر تگال، واقع در کشور اندونزی و در بازه زمانی ۱۹۸۳-۲۰۱۵ محاسبه کردند و شدیدترین خشکسالی ها را معین نمودند. نتایج این تحقیق بیانگر قابلیت بالای SPI در تعیین مدت و شدت خشکسالی بود. لی و همکاران (۲۰۲۰) به منظور پایش روند خشکسالی در کشور چین، SPI و شاخص تبخیر- تعرق و بارش استاندارد^۲ (SPEI) را برای بازه زمانی

2 Standardized Precipitation Evapotranspiration Index

۲۰۱۵-۱۹۸۰ در نظر گرفتند و نتایج مطالعه آنها نشان داد که رویداد های خشکسالی شناسایی شده با SPEI شدیدتر از SPI می باشد. رحمتی و همکاران (۲۰۲۰) در جنوب شرقی کوئینزلند استرالیا و طی دوره ۲۰۱۳-۱۹۹۴ با به کارگیری مدل جنگل تصادفی^۳ (RF)، ماشین بردار پشتیبان^۴ (SVM)، درختان رگرسیون تقویت شده^۵ (BRT)، درخت طبقه بندی و رگرسیون^۶ (CRT)، خطوط رگرسیونی تطبیقی چند متغیره^۷ (MARS) و آنالیز متمایز کننده انعطاف پذیر^۸ (FDA) اقدام به پیش بینی خشکسالی نمودند و طبق نتیجه اعتبار سنجی آن ها، مدل جنگل تصادفی بالاترین میزان دقت را دارا بود. شمشیربند و همکاران (۲۰۲۰) شاخص SPI، SPEI و شاخص جریان استاندارد (SSI)^۹ را به کمک روش رگرسیون بردار پشتیبان^{۱۰} (SVR)، برنامه نویسی بیان ژن^{۱۱} (GEP) و درختان مدل^{۱۲} (MT)، مدل سازی کردند و نتیجه تحقیق آنها دقت قابل قبول SPI بود. دیکشیت و همکاران (۲۰۲۰) به منظور تعیین روند خشکسالی منطقه ولز در استرالیا بازه زمانی ۱۹۰۱-۲۰۱۸ و SPEI را در نظر گرفتند. آن ها روش جنگل تصادفی را بر روی داده های مشاهداتی پیاده نمودند و در مرحله اعتبار سنجی به ضریب همبستگی بالا و میانگین مربعات ریشه مطلوب دست یافتند و به این نتیجه رسیدند که مدل جنگل تصادفی عملکرد بالایی را داراست. پارک و کیم (۲۰۱۹) با به کارگیری داده های توپوگرافی و تصاویر ماهواره ای، مدل پیش بینی ناحیه شدید خشکسالی را ارائه کردند و نیز مدل RF را هم به کار بردند و به دقت بالا در مدلسازی با RF دست یافتند.

شو و همکاران (۲۰۲۰)، SPI را در مقیاس زمانی ۱،۳،۶،۱۲ ماهه به کمک مدل میانگین متحرک خودهمبسته یکپارچه (ARIMA)^{۱۳} و مدل ARIMA-SVR را با یکدیگر مقایسه نمودند و در نهایت به این نتیجه رسیدند که در اعتبار

3 Random Forest

4 Support Vector Machine

5 Boosted Regression Trees

6 Classification And Regression Tree

7 Multivariate Adaptive Regression Splines

8 Flexible Discriminant Analysis

9 Standardized Streamflow Index

10 Support Vector Regression

11 Gene Expression Programming

12 Model Trees

13 Autoregressive Integrated Moving Average



شکل ۱. نقشه استان سیستان و بلوچستان و محدوده مطالعاتی زابل و زاهدان

شکل ۱. نقشه استان سیستان و بلوچستان و محدوده مطالعاتی زابل و زاهدان

شکل ۱. نقشه استان سیستان و بلوچستان و محدوده مطالعاتی زابل و زاهدان

شکل ۱. نقشه استان سیستان و بلوچستان و محدوده مطالعاتی زابل و زاهدان

مواد و روش‌ها

الف- منطقه مطالعاتی و داده‌ها: محدوده مطالعاتی ایستگاه‌های همدیدی شهرهای زابل و زاهدان، واقع در استان سیستان و بلوچستان می‌باشد که موقعیت آن در شکل (۱) آورده شده است. وسعت استان سیستان و بلوچستان ۱۸۱۴۷۱ کیلومتر مربع می‌باشد. این محدوده بین عرض جغرافیایی ۲۵° تا ۳۱° ۲۷ شمالی و طول جغرافیایی ۵۸° ۵۰ تا ۶۳° ۲۱ شرقی واقع شده است (موگبلی و همکاران، ۲۰۲۰). این منطقه دارای آب و هوای کویری و نیمه بیابانی است و متوسط بارندگی سالانه آن حدود ۱۱۰ میلی‌متر و متوسط دمای آن ۲۲ تا ۳۷ درجه سلسیوس است. رطوبت نسبی کم، کم بودن بارندگی سالانه نسبت به بارندگی متوسط کشور و تغییرات دمایی بالا، بخشی از ویژگی‌های این ناحیه می‌باشد (جوزگی و همکاران، ۲۰۱۸). یکی دیگر از ویژگی‌های مهم این منطقه، باد های ۱۲۰ روزه سیستان است که معمولاً از اواخر بهار تا اوایل پاییز ادامه دارد.

سنجی مدل‌ها بین داده‌های مشاهداتی و پیش‌بینی شده، دقت پیش‌بینی مدل ARIMA-SVR از مدل ARIMA بالاتر می‌باشد. استیسیو و همکاران (۲۰۲۱) زنجیره مارکوف را با به کارگیری SPI به منظور پیش‌بینی خشکسالی توسعه دادند و عدم قطعیت‌های آماری استفاده از زنجیره مارکوف در مورد پیش‌بینی خشکسالی سالانه را ارزیابی نمودند. آن‌ها برای ایجاد زنجیره مارکوف از ۱۰۰ سال داده بارش و برای اعتبارسنجی مدل خود از ۸ سال داده بارش استفاده کردند. به منظور ارزیابی کارایی مدل نیز از نسبت احتمال استفاده نمودند. سنجش عدم قطعیت‌ها نشان داد که مدل زنجیره مارکوف به منظور پیش‌بینی طبقه خشکسالی ممکن است در مقایسه با نسبت تاریخی طبقه خشکسالی از کارایی کمتری برخوردار باشد. انکراه و همکاران (۲۰۲۱) برای ساحل وینباً واقع در کشور غنا، به مقایسه SPI و SPEI طی بازه زمانی سال‌های ۲۰۱۹-۱۹۸۰ پرداختند. آن‌ها از ضریب همبستگی پیرسون و همبستگی متقابل و نمودار بلاند و آلتمن^{۱۵} به منظور مقایسه شدت، جهت و درجه توافق میان SPI و SPEI استفاده کردند. بر اساس نتایج مطالعه آن‌ها، بین SPI و SPEI در مقیاس ۶،۳،۱ و ۱۲ ماهه ارتباط مثبت و قوی به ترتیب معادل ۰،۹۰، ۰،۹۱، ۰،۸۴ و ۰،۹۳ وجود دارد. با توجه به مطالعات اخیر صورت گرفته، الگوریتم جنگل تصادفی و شاخص بارش استاندارد در ارزیابی خشکسالی از کارایی بالایی برخوردار می‌باشند. در مقالات پیشین، کمتر به بحث ارزیابی شدت خشکسالی پرداخته شده است لذا در این تحقیق شدت خشکسالی در طول سری زمانی محاسبه شده است و همچنین انجام تحلیل خودهمبستگی جزئی^{۱۶} (PACF) پیش از الگوریتم جنگل تصادفی به منظور بهبود تخمین SPI در نظر گرفته شده است. هدف تحقیق حاضر ارزیابی خشکسالی در منطقه مورد مطالعه در بازه زمانی ۳۰ ساله با به کارگیری شاخص بارش استاندارد، الگوریتم جنگل تصادفی و تحلیل PACF می‌باشد. پس از محاسبه SPI در دو مقیاس فصلی و سالانه برای هر دو ایستگاه، فراوانی وقوع طبقات خشکسالی نیز شدت خشکسالی محاسبه می‌گردد و پس از انجام تحلیل

14 Winneba-Ghana

15 Bland and Altman plot

16 Partial Autocorrelation Function

جدول ۱. مشخصات ایستگاه زابل و زاهدان (مرادی دشتیچردی و همکاران، ۲۰۱۵).

نام ایستگاه	طول جغرافیایی بر حسب درجه	عرض جغرافیایی بر حسب درجه	ارتفاع از سطح دریا بر حسب متر	متوسط بارش سالانه بر بر حسب میلی متر	نوع اقلیم
زابل	۶۱.۴۸	۳۱.۰۳	۴۸۹.۲	۶۲.۲	بسیار خشک
زاهدان	۶۰.۸۸	۲۹.۴۷	۱۳۷۰	۷۵.۳	خشک

جدول ۲. پارامترهای آماری سری زمانی بارش ماهانه در ایستگاه زابل و زاهدان

نام ایستگاه	انحراف معیار استاندارد	میانگین کل بارش های ماهانه در سری زمانی	مجموع کل بارش ماهانه در سری زمانی	حداکثر بارش ماهانه در سری زمانی	حداقل بارش ماهانه در سری زمانی	تعداد کل داده ها
زابل	۹.۵۱۲۸۴	۴.۴۵۷	۱۶۰۴.۵۱	۷۲.۲	۰	۳۶۰
زاهدان	۱۲.۴۷۶۰۹	۶.۵۳۱۲	۲۳۵۱.۲۴	۸۵.۴	۰	۳۶۰

سرعت باد های ۱۲۰ روزه نزدیک به ۱۰ متر بر ثانیه می باشد که موجب افزایش نرخ تیخیر، افزایش فرسایش خاک، کاهش پوشش گیاهی، طوفان های گرد و غبار و مشکلات تنفسی برای مردم می گردد (موگیلی و همکاران، ۲۰۲۰).

اطلاعات مربوط به ایستگاه های مورد مطالعه شامل طول جغرافیایی، عرض جغرافیایی، ارتفاع از سطح دریا، متوسط بارش سالانه و اقلیم در جدول (۱) ارائه شده است (مرادی دشتیچردی و همکاران، ۲۰۱۵).

داده های بارش ماهانه ایستگاه های هواشناسی شهرستان زابل و زاهدان در بازه زمانی دسامبر ۱۹۹۰ تا نوامبر ۲۰۲۰ به صورت سری زمانی از سازمان هواشناسی کشور اخذ و برای محاسبه و تحلیل های آماری ذکر شده در جدول (۲) از نرم افزار SPSS استفاده شد.

در این مطالعه برای اطمینان یافتن از صحت داده ها از آزمون همگن بودن داده ها به روش ران تست^{۱۷} و وضعیت روند در داده ها به روش تاوکندال^{۱۸} در محیط نرم افزار استفاده شد.

ب- شاخص بارش استاندارد (SPI): شاخص بارش استاندارد در سال ۱۹۹۳ توسط مک کی برای ارزیابی خشکسالی اقلیمی ابداع شد که داده های ماهانه بارندگی را در دوره مورد نظر ملاک قرار می دهد و توزیع آماری گاما را به عنوان تابع توزیع مناسب برای برازش روی داده های بارش استفاده می کند. این شاخص برای مقیاس های زمانی

مقدار شاخص به دست آمده برای داده های یک دوره زمانی طولانی، تابع توزیع نرمال می باشد که میانگین صفر و انحراف معیار ۱ دارد. بر اساس شاخص بارش استاندارد، زمانی که مقدار شاخص منفی گردد، خشکسالی اتفاق می افتد و هنگامی که مقدار شاخص مثبت گردد، خشکسالی پایان می پذیرد. برای محاسبه این شاخص حداقل ۳۰ سال دوره آماری لازم است (کمرزامن و همکاران، ۲۰۱۹). SPI با برازش دادن یک تابع چگالی احتمال به توزیع فراوانی بارش محاسبه می شود و این عمل را برای هر مقیاس زمانی دلخواه می توان به کار برد.

شهر تابع چگالی احتمال به یک توزیع نرمال استاندارد تبدیل می شود. تابع چگالی احتمال از رابطه (۱) قابل محاسبه است:

$$g(x) = \frac{1}{\beta \Gamma(\alpha)} x^{\alpha-1} e^{-x/\beta} \quad x > 0 \quad (1)$$

که $\alpha > 0$ و $\beta > 0$ به ترتیب فاکتورهای شکل و مقیاس هستند و همچنین $x > 0$ مقدار بارش و $\Gamma(\alpha)$ تابع گاما می باشد که از رابطه (۲) محاسبه می شود:

$$\Gamma(\alpha) = \int_0^{\infty} y^{\alpha-1} e^{-y} dy \quad (2)$$

مک کی و ادوردز بر اساس تقریب تام^{۱۹} به منظور حداکثر

17 Run Test
18 Kendall Tau

19 Approximation of Thom

تابع احتمال تجمعی به متغیر تصادفی نرمال استاندارد با میانگین صفر و واریانس یک تبدیل می‌شود که در واقع مقدار SPI می‌باشد. برای محاسبه مقادیر Z یا SPI از معادله پیشنهاد شده Stegun - Abramowitz به صورت روابط (۱۰) و (۱۱) و (۱۲) و (۱۳) استفاده شده است که این روابط، احتمالات تجمعی را به متغیر تصادفی نرمال استاندارد تبدیل می‌کنند، مقدار ضرایب ثابت معادلات فوق بیان شده است (مرادی دشتیچردی و همکاران، ۲۰۱۵). در جدول (۳) حدود SPI و طبقات خشکسالی ذکر شده است (شین و همکاران، ۲۰۱۸).

به ازای $0 < H(x) \leq 0.5$

$$t = \sqrt{\ln\left(\frac{1}{H(x)^2}\right)} \quad (10)$$

$$Z = \text{SPI} = -\left(t - \frac{c_0 + c_1 t + c_2 t^2}{1 + d_1 t + d_2 t^2 + d_3 t^3}\right) \quad (11)$$

و به ازای $0.5 < H(x) \leq 1$

$$t = \sqrt{\ln\left(\frac{1}{(1-H(x))^2}\right)} \quad (12)$$

$$Z = \text{SPI} = \left(t - \frac{c_0 + c_1 t + c_2 t^2}{1 + d_1 t + d_2 t^2 + d_3 t^3}\right) \quad (13)$$

$$(c_0 = 2.51551, c_1 = 0.802853, c_2 = 0.010328, \\ d_1 = 1.432788, d_2 = 0.189269, d_3 = 0.00130)$$

ج - تحلیل خود همبستگی جزئی: تابع خود همبستگی جزئی یک روش خطی می‌باشد که در شناسایی مدل های کاهنده خودکار استفاده می‌شود (سکیورا و همکاران، ۲۰۱۸). تابع خود همبستگی جزئی، همبستگی بین Z_t و Z_{t+k} را بعد از حذف متغیرهای همبسته $Z_{t-k+1}, \dots, Z_{t-2}, Z_{t-1}$ معین می‌کند. ضریب خود همبستگی جزئی معمولاً با Φ_{kk} نمایش داده می‌شود که در رابطه (۱۴) بیان شده است:

$$\Phi_{kk} = \text{Corr}(Z_t \text{ و } Z_{t-k} \mid Z_{t-1} \text{ و } Z_{t-2} \text{ و } \dots \text{ و } Z_{t-k+1}) \quad (14)$$

احتمال، ضریب های بهینه α و β را مطابق روابط (۳) و (۴) و (۵) بیان کردند:

$$\alpha = \frac{1}{4A} \left(1 + \sqrt{1 + \frac{4A}{3}}\right) \quad (3)$$

$$\beta = \frac{\bar{X}}{\alpha} \quad (4)$$

$$A = \text{Ln}(\bar{X}) - \frac{1}{n} \sum_{i=1}^n \text{Ln}(x_i) \quad (5)$$

در رابطه (۵)، n تعداد داده ها یا مشاهدات است. احتمال تجمعی بارش صورت گرفته در هر مقیاس زمانی از رابطه (۶) حاصل می‌گردد:

$$G(x) = \int_0^x g(x) dx = \frac{1}{\beta^\alpha \Gamma(\alpha)} \int_0^x x^{\alpha-1} e^{-x/\beta} dx \quad (6)$$

با جایگزین نمودن t به صورت $t = \frac{x}{\beta}$ تابع گامای جزئی از معادله (۷) محاسبه می‌شود:

$$G(x) = \frac{1}{\Gamma(\alpha)} \int_0^x t^{\alpha-1} e^{-t} dt \quad (7)$$

به این علت که تابع گاما برای X معادل صفر تعریف نشده است و در مناطق خشک و نیمه خشک ممکن است بارشی رخ ندهد، تابع احتمال تجمعی کل که شامل مقادیر صفر بارش باشد، به صورت رابطه (۸) تعریف می‌شود (مرادی دشتیچردی و همکاران، ۲۰۱۵).

$$H(x) = q + (1 - q) G(x) \quad (8)$$

q احتمال بارش معادل صفر است که از رابطه (۹) محاسبه می‌شود:

$$q = \frac{m}{n} \quad (9)$$

در رابطه (۹)، n تعداد کل داده ها و m تعداد داده های بارش دارای مقدار صفر در سری زمانی می‌باشد (مرادی و همکاران، ۲۰۱۱).

جدول ۳. طبقه بندی خشکسالی بر اساس SPI (شین و همکاران، ۲۰۱۸).

مقدار SPI	طبقه یا کلاس خشکسالی
> ۲	ترسالی بسیار شدید (Extremely wet)
۱.۵ تا ۱.۹۹	ترسالی شدید (Severely wet)
۱ تا ۱.۴۹	ترسالی متوسط (Moderately wet)
-۰.۹۹ تا ۰.۹۹	تقریباً نرمال (Near normal)
-۱ تا -۱.۴۹	خشکسالی متوسط (Moderately dry)
-۱.۵ تا -۱.۹۹	خشکسالی شدید (Severely dry)
< -۲	خشکسالی بسیار شدید (Extremely dry)

عصبی-فازی، SVM، SVR، دسته بندی، RF، رگرسیون لوجستیک و درختان تصمیم اشاره کرد (مسوی و همکاران، ۲۰۱۸، باربوزا و همکاران، ۲۰۱۷). الگوریتم جنگل تصادفی نخستین بار توسط Breiman در سال ۲۰۰۱ پیشنهاد شد. این الگوریتم برای بهبود پیش بینی، چندین درخت تصمیم تصادفی را ترکیب می‌کند و میانگین نتیجه هر درخت را به عنوان نتیجه نهایی در نظر می‌گیرد. الگوریتم جنگل تصادفی در رگرسیون و طبقه بندی قابل به کارگیری است. در مدل جنگل تصادفی هر درخت به کمک نمونه بوت استرپ^{۲۱} ایجاد می‌شود (بريمن، ۲۰۰۱). به کمک زیر مجموعه ای تصادفی از متغیرها یا فاکتورهای هر گره، تقسیم هر درخت انجام می‌گردد (کاتلر و همکاران، ۲۰۰۷). تعداد متغیرهای تصادفی در هر گره و تعداد درختان جنگل دو پارامتر مهم در الگوریتم جنگل تصادفی می‌باشند. متغیرهای تصادفی در هر گره از میان کل متغیرهای موجود (ρ)، انتخاب می‌شوند. به طور پیش فرض تعداد متغیرهای تصادفی در هر گره (m_{try}) در نرم افزارها در فرایند رگرسیون $\frac{\rho}{3}$ و در فرایند طبقه بندی $\sqrt{\rho}$ می‌باشد اما امکان تغییر آن وجود دارد (ایلیا و همکاران، ۲۰۱۸). در الگوریتم جنگل تصادفی از نمونه‌های خارج از کیسه^{۲۲} (OOB) به منظور ارزیابی مدل استفاده می‌گردد. OOB میانگین خطای پیش بینی در آموزش برای داده X_i است که فقط از درختان فاقد X_i در نمونه بوت استرپ استفاده می‌نماید.

$Y(x_i)$ میانگین پیش بینی‌های OOB است. تعداد درختان جنگل (n_{tree}) باید در حد کافی زیاد باشد تا خطای

Φ_{kk} ضریب همبستگی بین دو متغیر تصادفی Z_t و Z_{t-k} به شرط $Z_{t-1}, Z_{t-2}, \dots, Z_{t-k+1}$ می‌باشد. در محاسبه ضرایب خودهمبستگی جزئی معمولاً از معادله Yule-Walker که در رابطه (۱۵) آورده شده است، استفاده می‌شود:

$$\rho_1 = \Phi_{k1} + \Phi_{k2}\rho_1 + \dots + \Phi_{kk}\rho_{k-1} \quad (15)$$

$$\rho_2 = \Phi_{k1}\rho_1 + \Phi_{k2} + \dots + \Phi_{kk}\rho_{k-2}$$

تا

$$\rho_k = \Phi_{k1}\rho_{k-1} + \Phi_{k2}\rho_1 + \dots + \Phi_{kk}$$

ضرایب خودهمبستگی جزئی به کمک ضرایب خودهمبستگی جزئی با تغییر مقدار ρ در معادله Yule-Walker با r و شمارش k به صورت $k=1,2,\dots$ می‌تواند تخمین زده شود. مقدار Φ_{kk} با به کارگیری قوانین کرامر قابل محاسبه است (تننکی، ۲۰۱۹).

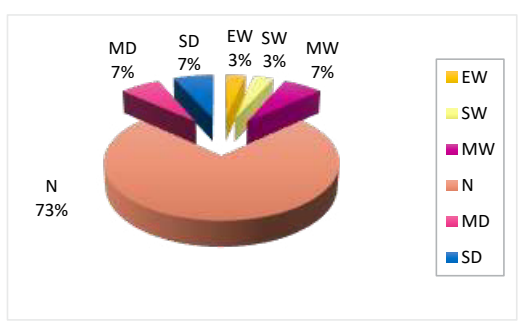
د - الگوریتم جنگل تصادفی (RF): اجرای آسان، امکان آموزش سریع، اعتبار سنجی و رسیدن به دقت بالا از ویژگیهای روش‌های یادگیری ماشین محسوب می‌شوند. همچنین به کمک این روش‌ها می‌توان سیستم‌های پیچیده و غیر خطی را با دقت بالا مدل‌سازی نمود (مسوی و همکاران، ۲۰۱۸). در روش‌های یادگیری ماشین، داده‌های ورودی آموزش داده می‌شود و از مدل آموزش دیده به منظور طبقه بندی، پیش بینی و یا خوشه بندی نمونه‌های آزمایشی استفاده می‌شود (لیاکس و همکاران، ۲۰۱۸). در مورد انواع مدل‌های یادگیری ماشین می‌توان به شبکه‌های عصبی مصنوعی (ANN)^{۲۰}،

21 Bootstrap Sample

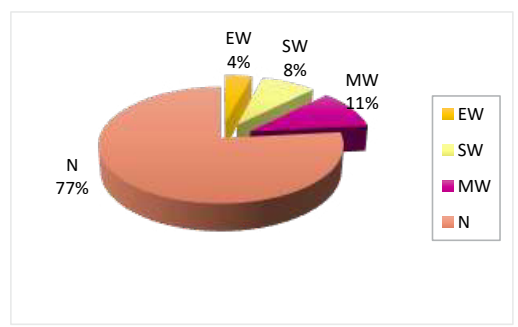
22 Out of Bag

20 Artificial Neural Network

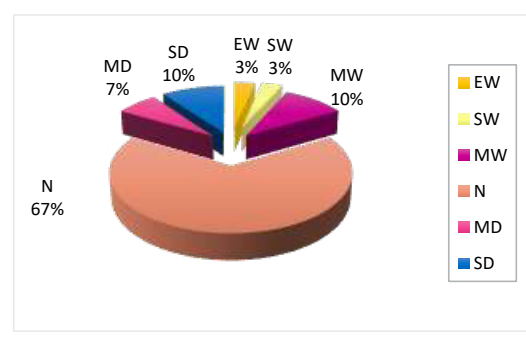
زیب تری و همکاران



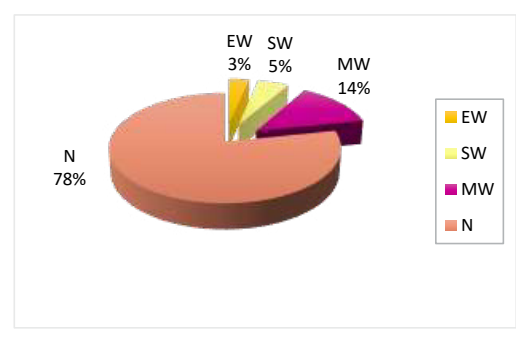
شکل ۳. درصد رخدادهای طبقات مختلف خشکسالی SPI12 در زابل



شکل ۲. درصد رخدادهای طبقات مختلف خشکسالی SPI3 در زابل



شکل ۵. درصد رخدادهای طبقات مختلف خشکسالی SPI12 در زاهدان



شکل ۴. درصد رخدادهای طبقات مختلف خشکسالی SPI3 در زاهدان

(MAE) و شاخص توافقی^{۲۶} (IOA) سنجیده شد.

نمونه های OOB پایدار گردد. میزان خطای مربع میانگین از رابطه (۱۶) قابل محاسبه است (نوروزی و همکاران، ۲۰۲۰).

نتایج و بحث

همگن بودن داده های بارش به کمک آزمون ران تست و عدم وجود روند در داده های بارش به کمک آزمون تاوکندال در محیط نرم افزار SPSS بررسی گردید که بر اساس آن، داده های بارش هر دو ایستگاه همگن بوده و فاقد روند هستند. پارامتر های تابع توزیع گاما با استفاده از نرم افزار Easy 5.5 Fit تعیین شده و پنجره زمانی ثابت برای محاسبه SPI در نظر گرفته شد. به کمک نرم افزار Excel، SPI در مقیاس سالانه و فصلی برای ایستگاه های هواشناسی همدیدی زابل و زاهدان محاسبه شده است. با توجه به حدود SPI در سری زمانی مطابق جدول (۳)، درصد رخدادهای طبقات مختلف خشکسالی سالانه و فصلی در قالب نمودار دایره ای در شکل های (۲) تا (۵) ارائه شده است و نیز وضعیت SPI در فصل های بهار، تابستان، پاییز، زمستان و مقیاس سالانه به صورت سری زمانی در قالب نمودار خطی در دو ایستگاه در شکل های (۶) تا (۱۰)

$$MSE \cong MSE^{OOB} = \frac{1}{N} \sum_{n=1}^N (Y(x_i) - Y_i)^2 \quad (16)$$

مزیت مدل جنگل تصادفی شامل واریانس پایین، کم بودن همبستگی میان درخت ها، دقت بالا در پیش بینی و تخمین خطا با استفاده از داده خارج از کیسه، عدم نیاز به هرس کردن درختان، توانایی تحلیل روابط غیرخطی و اندازه گیری اهمیت پارامتر ها می باشد (اورینگم و همکاران، ۲۰۱۶، نوروزی و همکاران، ۲۰۲۰).

ه - اعتبارسنجی مدل سازی: اعتبار مدل سازی جنگل تصادفی به کمک پارامتر های آماری ضریب همبستگی (R)، خطای جذر میانگین مربعات^{۲۳} (RMSE)، نرمال شده خطای جذر میانگین مربعات^{۲۴} (NRMSE)، میانگین خطای مطلق^{۲۵}

23 Root Mean Square Error
24 Normalized Root Mean Square Error
25 Mean Absolute Error

26 Index Of Agreement

در سری زمانی حضور دارند با یکدیگر جمع شده و بر مدت خشکسالی تقسیم شده اند.

بر اساس شکل های (۲) تا (۵) در بازه ۳۰ ساله مورد بررسی در مقیاس فصلی SPI، طبقه خشکسالی تقریباً

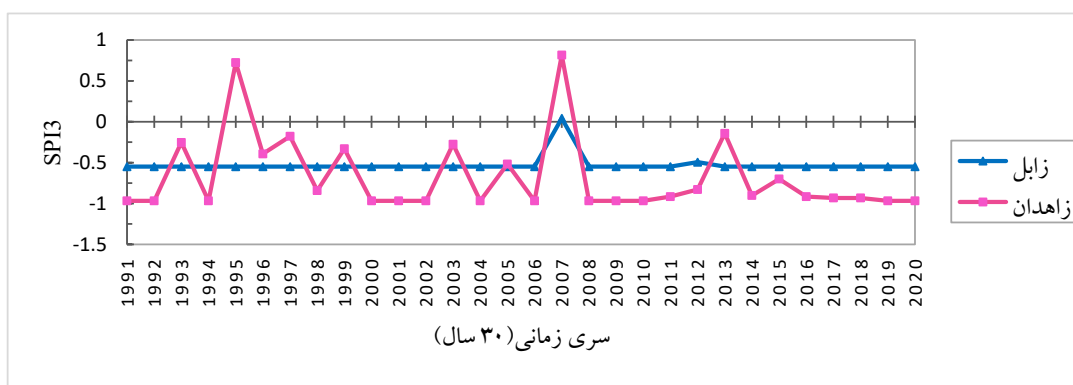
نمایش داده شده است. کمترین مقدار SPI، تاریخ وقوع اوج خشکسالی و شدت خشکسالی در طول سری زمانی در ایستگاه زابل و زاهدان در جدول (۴) مشخص شده است. برای تعیین شدت خشکسالی مقادیر منفی SPI که به صورت متوالی



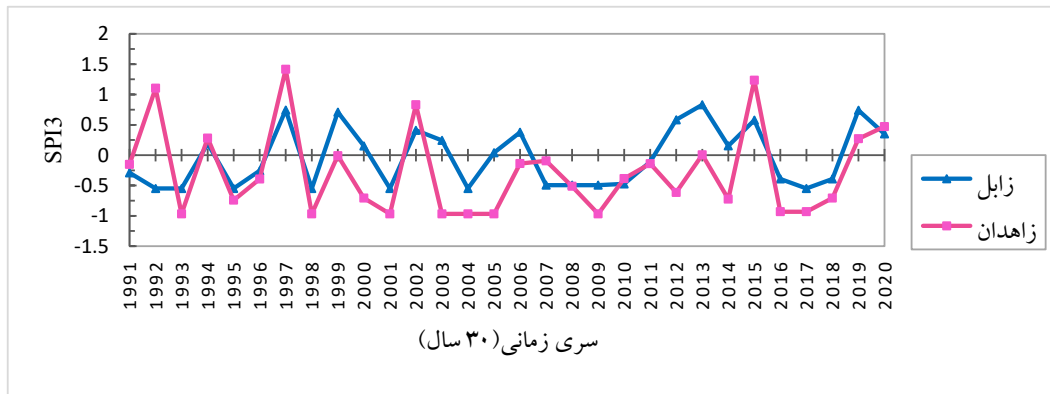
شکل ۶. نمودار SPI ایستگاه زابل و زاهدان در فصل زمستان



شکل ۷. نمودار SPI ایستگاه زابل و زاهدان در فصل بهار



شکل ۸. نمودار SPI ایستگاه زابل و زاهدان در فصل تابستان



شکل ۹. نمودار SPI ایستگاه زابل و زاهدان در فصل پاییز



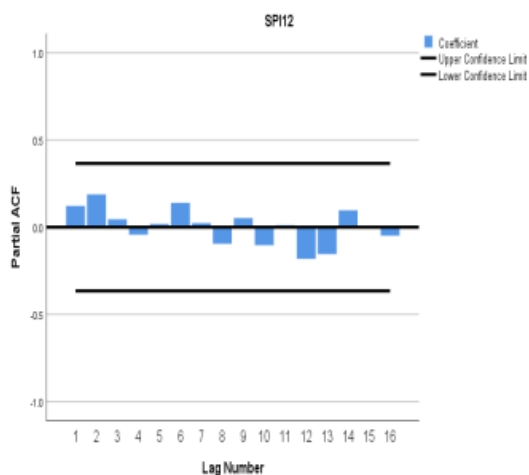
شکل ۱۰. نمودار SPI ایستگاه زابل و زاهدان در مقیاس سالانه

جدول ۴. کمترین مقدار SPI، تاریخ وقوع اوج خشکسالی و شدت خشکسالی در طول سری زمانی در ایستگاه زابل و زاهدان

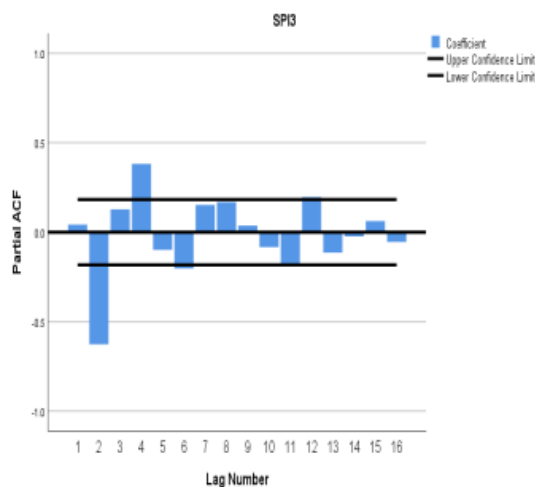
نام ایستگاه	مقیاس مورد بررسی	کمترین مقدار SPI در سری زمانی	تاریخ وقوع اوج خشکسالی	شدت خشکسالی در طول سری زمانی
زابل	فصلی	-۰.۵۴۸۵۲۲۸۳	فصل های فاقد بارش	-۱۴.۹۰۶۰۶۹۰۳
زابل	سالانه	-۱.۸۱۱۷۵۴۳۴۹	سال ۲۰۰۱	-۲.۳۰۷۴۰۲۵۸۸
زاهدان	فصلی	-۰.۹۶۷۴۲۱۵۶۶	فصل های فاقد بارش	-۲۰.۰۳۴۴۹۵۹۳
زاهدان	سالانه	-۱.۹۸۶۶۴۰۶۵۹	سال ۲۰۰۱	-۴.۶۹۴۳۱۹۲۶۵

تاخیر زمانی هایی که بالاتر یا پایین تر از خطوط معناداری (خطوط سیاه رنگ) هستند، معنادار بوده و به عنوان ورودی الگوریتم RF استفاده شده اند. بر اساس تحلیل PACF چنین استنباط می شود که در مقیاس سالانه SPI به دلیل مسائلی چون تغییر اقلیم در سال های مختلف وابستگی خاصی میان SPI ها در دو ایستگاه وجود ندارد در حالی که در مقیاس فصلی SPI، میان SPI

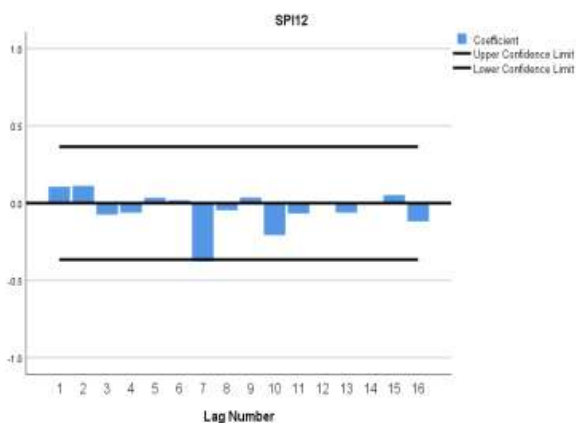
نرمال (\bar{N}) بیشترین فراوانی وقوع را برای ایستگاه زابل و زاهدان دارد که به ترتیب ۷۷ درصد و ۷۸ درصد می باشد همچنین در مقیاس سالانه SPI، طبقه خشکسالی تقریباً نرمال (\bar{N}) بیشترین فراوانی وقوع را برای ایستگاه زابل و زاهدان دارد که به ترتیب ۷۳ درصد و ۶۷ درصد می باشد. نتایج تحلیل PACF برای SPI فصلی و سالانه در نرم افزار SPSS در شکل های (۱۱) تا (۱۴) قابل مشاهده است.



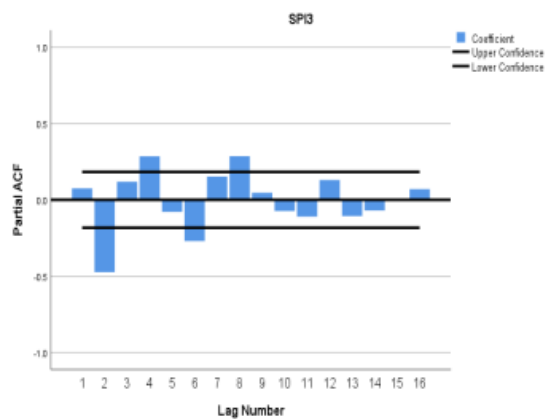
شکل ۱۲. نمودار تحلیل PACF در ایستگاه زابل برای SPI12



شکل ۱۱. نمودار تحلیل PACF در ایستگاه زابل برای SPI3



شکل ۱۴. نمودار تحلیل PACF در ایستگاه زاهدان برای SPI12



شکل ۱۳. نمودار تحلیل PACF در ایستگاه زاهدان برای SPI3

و بار دیگر تاخیر زمانی های ۶،۴،۲ و ۸ ماهه به عنوان ورودی الگوریتم جنگل تصادفی به کار رفته اند. ۷۰ درصد نمونه ها برای آموزش مدل و ۳۰ درصد نمونه ها در فرایند آزمایش استفاده شده اند. برای همگرا شدن خطا و افزایش دقت مدل، تعداد درختان جنگل (n_{tree})، ۵۰۰ درخت در نظر گرفته شده و متناسب با فرایند رگرسیون و تعداد ورودی ها، تعداد متغیر های تصادفی در هر گره (m_{try}) در نرم افزار Rstudio تنظیم شده که جزئیات آن در جدول (۵) ارائه شده است. با نوشتن کد نرم افزاری در محیط نرم افزار Rstudio، SPI فصلی، پیش بینی گردید و در شکل های (۱۶) تا (۱۹) مقادیر مشاهداتی و پیش بینی شده به صورت سری زمانی نمایش

ها همبستگی وجود دارد و از شاخص های دارای همبستگی به عنوان ورودی الگوریتم جنگل تصادفی استفاده گردید. بر اساس شکل (۱۱) که نمودار تحلیل PACF در ایستگاه زابل برای SPI3 می باشد، تاخیر زمانی های ۶،۴،۲ و ۱۲ ماهه معنادار هستند به همین دلیل به منظور مقایسه نمودن نتایج، یک بار تاخیر زمانی های ۶،۴،۲ و ۶ ماهه و بار دیگر تاخیر زمانی های ۶،۴،۲ و ۱۲ ماهه به عنوان ورودی الگوریتم جنگل تصادفی استفاده شده اند. همچنین بر اساس شکل (۱۳) که نمودار تحلیل PACF در ایستگاه زاهدان برای SPI3 می باشد، تاخیر زمانی های ۶،۴،۲ و ۸ ماهه معنادار هستند به همین دلیل به منظور مقایسه نمودن نتایج، یک بار تاخیر زمانی های ۶،۴،۲ و ۶ ماهه

نتیجه هر درخت، SPI در الگوریتم جنگل تصادفی تخمین زده می‌شود.

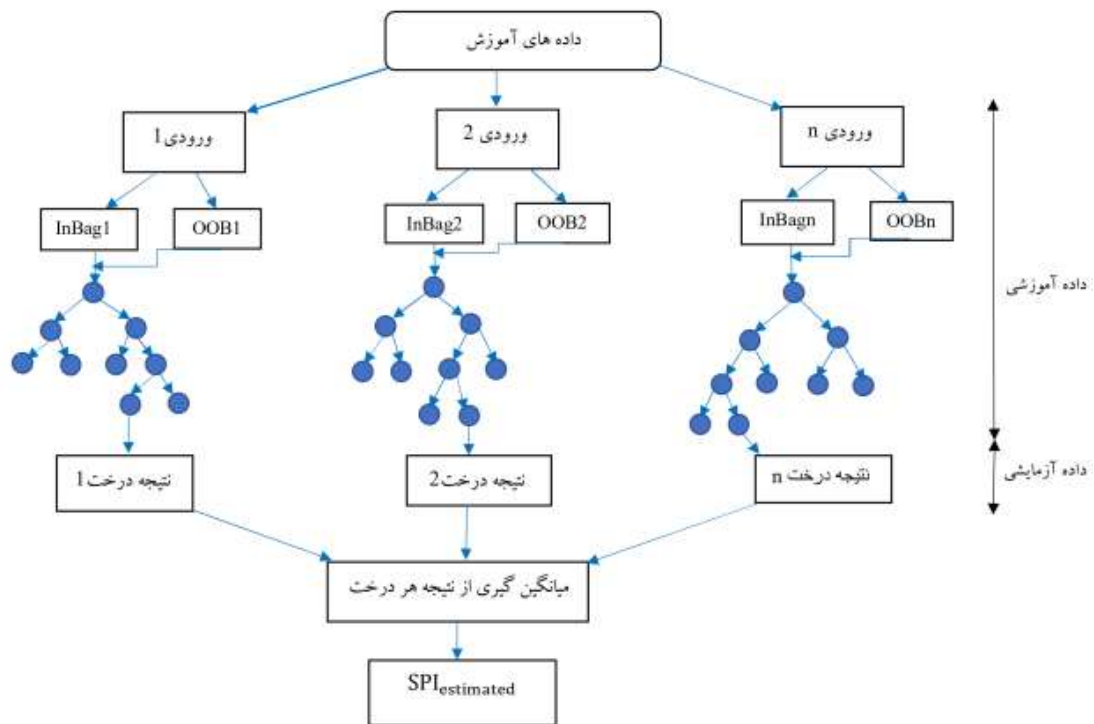
پس از تخمین SPI به کمک الگوریتم جنگل تصادفی، به منظور بررسی اعتبار و دقت الگوریتم جنگل تصادفی، بین مقادیر مشاهداتی SPI و مقادیر پیش بینی شده SPI در دو ایستگاه در مورد مجموعه آموزشی و آزمایشی پارامترهای آماری R ، $RMSE$ ، $NRMSE$ ، MAE و IOA مورد استفاده قرار گرفتند که نتایج آن در جداول (۶) و (۷) ارائه شده است. مقادیر بیشتر پارامترهای R و IOA و در عین حال مقادیر کمتر پارامترهای MAE ، $NRMSE$ ، $RMSE$

داده شده است که اعداد ۳،۲،۱ و ۴ در نمودارها به ترتیب بیانگر فصل‌های زمستان، بهار، تابستان و پاییز می‌باشند.

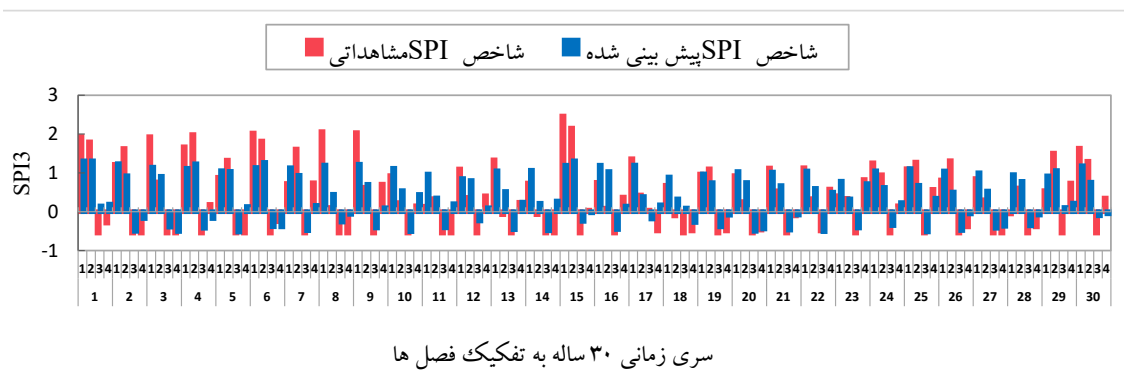
در شکل (۱۵) شماتیک مکانیسم روش جنگل تصادفی جهت تخمین SPI نشان داده شده است که هر درخت به کمک نمونه بوت استرپ ایجاد می‌شود و به کمک زیرمجموعه ای تصادفی از متغیرها یا فاکتورهای هر گره، تقسیم هر درخت انجام می‌شود. متغیرهای تصادفی در هر گره از میان کل متغیرهای موجود (ρ)، انتخاب می‌شوند و از نمونه‌های خارج از کیسه (OOB) به منظور ارزیابی مدل استفاده می‌گردد. پس از تقسیم شدن هر درخت، با میانگین‌گیری از

جدول ۵. تنظیمات الگوریتم جنگل تصادفی

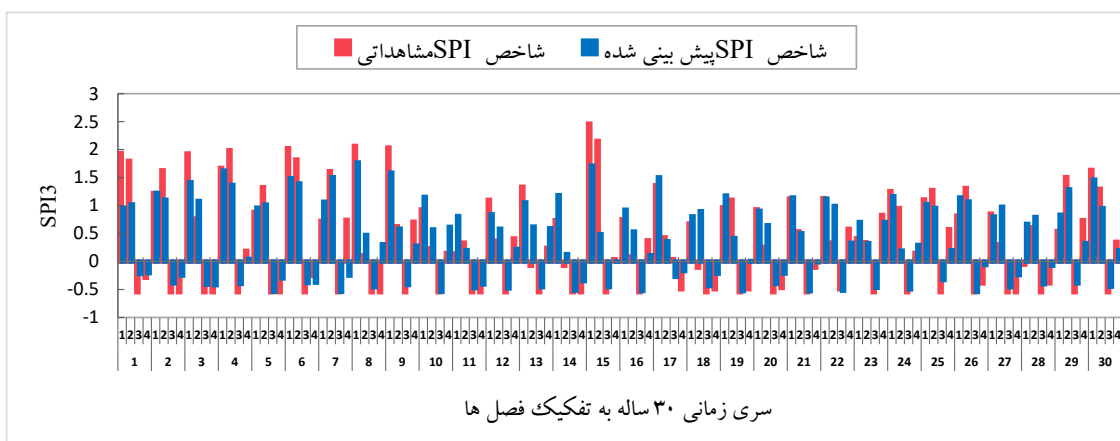
نام ایستگاه	مقیاس SPI	ورودی‌های مدل جنگل تصادفی	درصد نمونه‌های آزمایش	درصد نمونه‌های آموزش	mtry	ntree
زابل	فصلی	تاخیر زمانی ۲،۴۶،۱۲	۳۰	۷۰	۲	۵۰۰
		تاخیر زمانی ۲،۴۶	۳۰	۷۰	۱	۵۰۰
زاهدان	فصلی	تاخیر زمانی ۲،۴۶،۸	۳۰	۷۰	۲	۵۰۰
		تاخیر زمانی ۲،۴۶	۳۰	۷۰	۱	۵۰۰



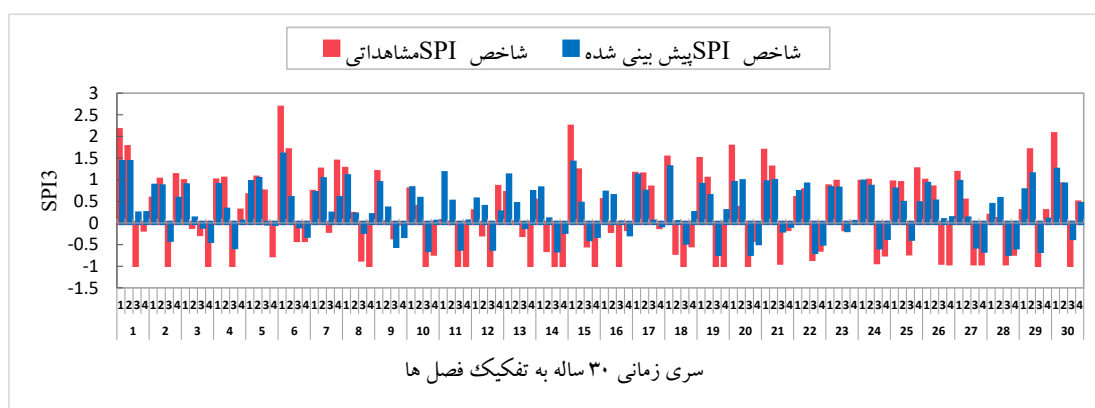
شکل ۱۵. نمایشی از مکانیسم روش جنگل تصادفی برای تخمین SPI



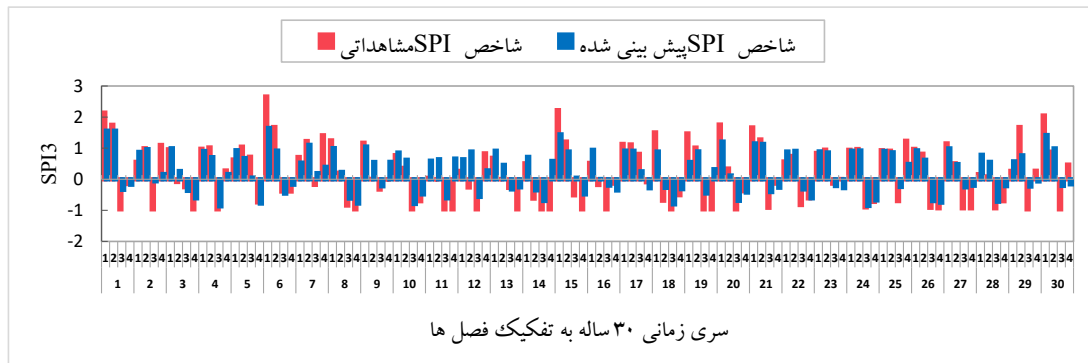
شکل ۱۶. نمودار مقادیر مشاهداتی و پیش بینی شده SPI3 ایستگاه زابل در RF با ورودی تاخیر زمانی



شکل ۱۷. نمودار مقادیر مشاهداتی و پیش بینی شده SPI3 ایستگاه زابل در RF با ورودی تاخیر زمانی 2,4,6,12



شکل ۱۸. نمودار مقادیر مشاهداتی و پیش بینی شده SPI3 ایستگاه زاهدان در RF با ورودی تاخیر زمانی 2,4,6



شکل ۱۹. نمودار مقادیر مشاهداتی و پیش بینی شده SPI3 ایستگاه زاهدان در RF با ورودی تاخیر زمانی 2,4, 6,8

جدول ۶. اعتبارسنجی RF در مدلسازی SPI فصلی ایستگاه زابل در مجموعه داده های آموزشی و آزمایشی

ایستگاه	ورودی	مجموعه مورد بررسی	پارامتر آماری	مقدار پارامتر آماری
زابل	تاخیر زمانی ۲,۴,۶,۱۲	آموزش	R	۰.۹۴۵۷
			RMSE	۰.۲۸۷۸
			NRMSE	۹.۵۵۶۹
			MAE	۰.۲۱۶۴
			IOA	۰.۹۶۴۸
زابل	تاخیر زمانی ۲,۴,۶,۱۲	آزمایش	R	۰.۷۷۰۶
			RMSE	۰.۵۴۳۱
			NRMSE	۲۰.۰۹۱۵
			MAE	۰.۴۱۷۵
			IOA	۰.۸۵۵۶
زابل	تاخیر زمانی ۲,۴,۶	آموزش	R	۰.۸۹۱۲
			RMSE	۰.۳۹۰۷
			NRMSE	۱۴.۴۵۲۶
			MAE	۰.۳۰۲۸
			IOA	۰.۹۲۵۶
زابل	تاخیر زمانی ۲,۴,۶	آزمایش	R	۰.۸۴۶۴
			RMSE	۰.۵۰۷
			NRMSE	۱۶.۸۳۴۷
			MAE	۰.۴۰۳۳
			IOA	۰.۸۶۷۳

تاخیر زمانی ۲,۴,۶ ماهه در مجموعه آزمایش دقت بالاتری حاصل شده است. پارامتر های آماری بررسی شده در جدول (۷) در مجموعه آموزش و آزمایش به ازای مقادیر ورودی تاخیر زمانی ۲,۴,۶,۸ ماهه نسبت به مقادیر ورودی تاخیر زمانی ۲,۴,۶ ماهه از دقت بیشتری برخوردار است. در ایستگاه زابل پارامتر آماری R برای مجموعه آموزشی برای هر یک از ورودی های مدل به ترتیب ۰/۹۵ و ۰/۸۹ است و برای مجموعه

نشان دهنده دقت مطلوب در تخمین SPI می باشد. در ادامه نتایج اعتبارسنجی الگوریتم جنگل تصادفی برای هر ایستگاه، به ازای مقادیر ورودی متفاوت از تاخیر زمانی و در نظر گرفتن مجموعه آموزش و آزمایش مقایسه می شوند. بر اساس جدول (۶) پارامتر های آماری بررسی شده به ازای مقادیر ورودی تاخیر زمانی ۲,۴,۶,۱۲ ماهه در مجموعه آموزش از دقت بیشتری برخوردار بوده در حالی که به ازای مقادیر ورودی

جدول ۷. اعتبارسنجی RF در مدلسازی SPI فصلی ایستگاه زاهدان در مجموعه داده های آموزشی و آزمایشی

ایستگاه	ورودی	مجموعه مورد بررسی	پارامتر آماری	مقدار پارامتر آماری
زاهدان	تاخیر زمانی ۲,۴۶,۸	آموزش	R	۰.۹۳۴۹
			RMSE	۰.۳۶۳
			NRMSE	۱۰.۰۰۵۷
			MAE	۰.۲۸۲
			IOA	۰.۹۴۹۵
زاهدان	تاخیر زمانی ۲,۴۶,۸	آزمایش	R	۰.۷۰۱۳
			RMSE	۰.۷۱۳۹
			NRMSE	۲۶.۹۸۲۴
			MAE	۰.۵۹۸۲
			IOA	۰.۷۴۰۸
زاهدان	تاخیر زمانی ۲,۴,۶	آموزش	R	۰.۹۱۶۶
			RMSE	۰.۴۳۲۵
			NRMSE	۱۱.۹۲۲۶
			MAE	۰.۳۴۳۳
			IOA	۰.۹۲۰۵
زاهدان	تاخیر زمانی ۲,۴,۶	آزمایش	R	۰.۶۳۲۵
			RMSE	۰.۷۴۰۶
			NRMSE	۲۷.۹۹۱۸
			MAE	۰.۶۴۰۲
			IOA	۰.۶۳۰۳

اجتماع، اقتصاد و محیط زیست می‌گردد. با تجزیه و تحلیل دقیق خشکسالی می‌توان برنامه های جامعی برای کاهش اثرات نامطلوب خشکسالی ارائه نمود. به دلیل رطوبت نسبی کم، کم بودن بارندگی سالانه نسبت به بارندگی متوسط کشور و تغییرات دمایی بالا، وجود باد های ۱۲۰ روزه و اثرات نامطلوب این باد ها روی نرخ تبخیر، فرسایش خاک، پوشش گیاهی، طوفان های گرد و غبار و ایجاد مشکلات تنفسی برای افراد، ایستگاه هواشناسی همدیدی زابل و زاهدان واقع در جنوب شرق کشور برای این مطالعه انتخاب شده است. با توجه به اهمیت موضوع در این تحقیق با محاسبه SPI در مقیاس فصلی و سالانه، طبقات خشکسالی، شدت خشکسالی و شدیدترین خشکسالی در طول سری زمانی ۳۰ ساله در ایستگاه زابل و زاهدان تعیین شده است. با توجه به نتایج به دست آمده می‌توان بیان داشت که در بازه ۳۰ ساله مورد بررسی، وضعیت غالب خشکسالی در مقیاس فصلی و سالانه برای ایستگاه زابل و زاهدان تقریباً نرمال

آزمایش به ترتیب ۰/۷۷ و ۰/۸۵ می‌باشد. در مورد ایستگاه زاهدان نیز پارامتر آماری R برای مجموعه آموزشی برای هر یک از ورودی های مدل به ترتیب ۰/۹۴ و ۰/۹۲ است و برای مجموعه آزمایش به ترتیب ۰/۷۰ و ۰/۶۳ می‌باشد. در ایستگاه زابل پارامتر آماری IOA برای مجموعه آموزشی برای هر یک از ورودی های مدل به ترتیب ۰/۹۷ و ۰/۹۳ است و برای مجموعه آزمایش به ترتیب ۰/۸۶ و ۰/۸۷ می‌باشد. در مورد ایستگاه زاهدان نیز پارامتر آماری IOA برای مجموعه آموزشی برای هر یک از ورودی های مدل به ترتیب ۰/۹۵ و ۰/۹۲ است و برای مجموعه آزمایش به ترتیب ۰/۷۴ و ۰/۶۳ می‌باشد. بر اساس سایر پارامتر های آماری بررسی شده در این تحقیق یعنی RMSE، NRMSE و MAE الگوریتم جنگل تصادفی در تخمین SPI از دقت نسبتاً بالایی برخوردار می‌باشد.

جمع‌بندی

خشکسالی باعث بروز خسارت های جبران ناپذیر به

- Tools in Coastal Winneba-Ghana]. *Journal of Geography, Environment and Earth Science*, 25(6).
- Barboza, F., Kimura, H., & Altman, E. 2017. Machine learning models and bankruptcy prediction. *Expert Systems with Applications*, 83, 405-417.
- Breiman, L. 2001. Random forests. *Machine Learning*, 45(1), 5-32.
- Chen, S. T., Kuo, C. C., & Yu, P. S. 2009. Historical trends and variability of meteorological droughts in Taiwan. *Hydrological sciences journal*, 54(3), 430-441.
- Choi, M., Jacobs, J. M., Anderson, M. C., & Bosch, D. D. 2013. Evaluation of drought indices via remotely sensed data with hydrological variables. *Journal of Hydrology*, 476, 265-273.
- Cutler, D.R., Edwards, T.C., Beard, K.H., Cutler, A., Hess, K.T., Gibson, J., Lawler, J.J. 2007. Random forests for classification in ecology. *Ecology* 88 (11), 2783 -2792.
- Dashtpajardi, M. M., Kousari, M. R., Vagharfard, H., Ghonchepour, D., Hosseini, M. E., & Ahani, H. 2015. An investigation of drought magnitude trend during 1975-2005 in arid and semi-arid regions of Iran. *Environmental earth sciences*, 73(3), 1231-1244.
- 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.
- Estácio, A. B. S., da Silva, S. M. O., & de Assis Souza Filho, F. 2021. STATISTICAL UNCERTAINTY IN DROUGHT FORECASTING USING MARKOV CHAINS AND THE STANDARD PRECIPITATION INDEX (SPI). *Revista Brasileira de Climatologia*, 28.
- Everingham, Y., Sexton, J., Skocaj, D., & Inman-Bamber, G. 2016. Accurate prediction of sugarcane yield using a random forest algorithm. *Agronomy for sustainable development*, 36(2), 27.
- Heim Jr, R. R. 2002. A review of twentieth-century drought indices used in the United States. *Bulletin of the American Meteorological Society*, 83(8), 1149-1166.
- Ilija, I., Loupasakis, C., & Tsangaratos, P. 2018. Land subsidence phenomena investigated by spatiotemporal analysis of groundwater resources, remote sensing techniques, and random forest method: the case of Western Thessaly, Greece. *Environmental monitoring and assessment*, 190(11), 1-19.
- Jozaghi, A., Alizadeh, B., Hatami, M., Flood, I., Khorrami, M., Khodaei, N., & Ghasemi Tousi, E. 2018. A comparative study of the AHP and TOPSIS techniques for dam site selection using GIS: a case study of Sistan and Baluchestan Province, Iran. *Geosciences*, 8(12), 494.
- Kamruzzaman, M., Cho, J., Jang, M.-W., & Hwang, S. 2019. Comparative evaluation of standardized precipitation index (SPI) and effective drought index (EDI) for meteorological drought detection over

می‌باشد. مقدار و روند تغییرات SPI در فصل زمستان، بهار، تابستان، پاییز و سالانه در ایستگاه زابل و زاهدان به صورت سری زمانی قابل مقایسه می‌باشد. شدت خشکسالی در ایستگاه زاهدان در مقیاس فصلی و سالانه نسبت به ایستگاه زابل بیشتر است و اوج خشکسالی سالانه هر دو ایستگاه در سال ۲۰۰۱ رخ داده است، همچنین اوج خشکسالی هر دو ایستگاه در مقیاس فصلی در فصولی رخ داده است که فاقد بارش بوده اند. بر اساس مطالب ذکر شده با بهره گیری از SPI می‌توان طبقات خشکسالی، روند تغییرات مقدار SPI در مقیاس فصلی و سالانه و شدت خشکسالی را ارزیابی و پیش نمود. با انجام تحلیل خودهمبستگی جزئی (PACF) در مورد SPI، ترکیبی از تاخیر زمانی های معنادار حاصل شده از تحلیل PACF به عنوان ورودی های الگوریتم جنگل تصادفی استفاده شده است. با اجرای الگوریتم جنگل تصادفی، مقادیر SPI تخمین زده شده و مقادیر مشاهداتی SPI و مقادیر تخمین زده شده SPI در دو ایستگاه به صورت سری زمانی ارائه شده اند. بر اساس اعتبارسنجی الگوریتم جنگل تصادفی به منظور تخمین SPI (برای دو مجموعه آموزشی و آزمایشی) به کمک پارامتر های آماری R، RMSE، NRMSE، MAE و IOA این الگوریتم دارای دقت قابل قبول و مطلوب می‌باشد. بر اساس نتایج اعتبارسنجی الگوریتم جنگل تصادفی در ایستگاه زابل در مورد مجموعه داده های آزمایشی، تاخیر زمانی های ۲،۴،۶ ماهه به عنوان ورودی الگوریتم، بالاترین میزان دقت را داشته است ($R=0/85$ ، $RMSE=0/51$ ، $NRMSE=16/83$ ، $MAE=0/40$ ، $IOA=0/87$) و همچنین در ایستگاه زاهدان با در نظر گرفتن مجموعه آزمایشی، تاخیر زمانی های ۲،۴،۶،۸ ماهه به عنوان ورودی الگوریتم، دقت بیشتری داشته است ($R=0/70$ ، $RMSE=0/71$ ، $NRMSE=26/98$ ، $MAE=0/60$ و $IOA=0/74$ ---). در نتیجه استفاده از تحلیل PACF در راستای تخمین میزان خشکسالی می‌تواند موثر باشد.

مراجع

- Ankrah, J. 2021. Comparative Assessment of Standard Precipitation Index and Standard Precipitation Evapotranspiration Index as Drought Evaluation

- 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.
- Shamshirband, S., Hashemi, S., Salimi, H., Samadianfard, S., Asadi, E., Shadkani, S., . . . Chau, K.-W. 2020. Predicting standardized streamflow index for hydrological drought using machine learning models. *Engineering Applications of Computational Fluid Mechanics*, 14(1), 339-350.
- Sharma, A., Dadhwal, V., Jeganathan, C., & Tolpekin, V. 2009. Drought monitoring using standardized precipitation index: a case study for the state of Karnataka, India. *Geospatial Application Papers: Environment-Forest*.
- Shin, J. Y., Chen, S., Lee, J.-H., & Kim, T.-W. 2018. Investigation of drought propagation in South Korea using drought index and conditional probability. *Terrestrial, Atmospheric & Oceanic Sciences*, 29(2).
- Siqueira, H., Boccato, L., Luna, I., Attux, R., & Lyra, C. 2018. Performance analysis of unorganized machines in streamflow forecasting of Brazilian plants. *Applied Soft Computing*, 68, 494-506.
- Tinungki, G. 2019. The analysis of partial autocorrelation function in predicting maximum wind speed. Paper presented at the IOP Conference Series: Earth and Environmental Science.
- Xu, D., Zhang, Q., Ding, Y., & Huang, H. 2020. Application of a Hybrid ARIMA-SVR Model Based on the SPI for the Forecast of Drought—A Case Study in Henan Province, China. *Journal of Applied Meteorology and Climatology*, 59(7), 1239-1259.
- Zandalinas, S. I., Fritschi, F. B., & Mittler, R. 2021. Global warming, climate change, and environmental pollution: Recipe for a multifactorial stress combination disaster. *Trends in Plant Science*. Bangladesh. *한국농공학회논문집*, 61(1), 145.
- Liakos, K. G., Busato, P., Moshou, D., Pearson, S., & Bochtis, D. 2018. Machine learning in agriculture: A review. *Sensors*, 18(8), 2674.
- Li, L., She, D., Zheng, H., Lin, P., & Yang, Z.-L. 2020. Elucidating diverse drought characteristics from two meteorological drought indices (SPI and SPEI) in China. *Journal of Hydrometeorology*, 21(7), 1513-1530.
- Li, Y., Xie, Z., Qin, Y., Xia, H., Zheng, Z., Zhang, L., ... & Liu, Z. 2019. Drought under global warming and climate change: An empirical study of the Loess Plateau. *Sustainability*, 11(5), 12811
- Moghbeli, A., Delbari, M., & Amiri, M. 2020. Application of a standardized precipitation index for mapping drought severity in an arid climate region, southeastern Iran. *Arabian Journal of Geosciences*, 13(5), 1-16.
- Moradi, H., Rajabi, M., & Faragzadeh, M. 2011. Investigation of meteorological drought characteristics in Fars province, Iran. *Catena*, 84(1-2), 35-46.
- Mosavi, A., Ozturk, P., & Chau, K.-w. 2018. Flood prediction using machine learning models: Literature review. *Water*, 10(11), 1536.
- Norouzi, H., & Moghaddam, A. A. 2020. Groundwater quality assessment using random forest method based on groundwater quality indices (case study: Miandoab plain aquifer, NW of Iran). *Arabian Journal of Geosciences*, 13(18), 1-13.
- Park, H., & Kim, K. 2019. Prediction of severe drought area based on random forest: Using satellite image and topography data. *Water*, 11(4), 705.
- Pramudya, Y., & Onishi, T. 2018. Assessment of the standardized precipitation index (SPI) in Tegal City, Central Java, Indonesia. Paper presented at the IOP Conference Series: Earth and Environmental Science.
- Rahmati, O., Falah, F., Dayal, K. S., Deo, R. C.,

Drought assessment using standardized precipitation index and random forest algorithm

Zeinab Torabi¹, Ramazan Vaghee^{2*}, Alireza Ghaemi³

¹ MSc Student in Civil and Environmental Engineering, Shahrood University of Technology, Shahrood, Iran.

² Assistant Professor in Civil and Environmental Engineering, Shahrood University of Technology, Shahrood, Iran.

³ Ph.D. Student in Civil Engineering, Faculty of water resource management, University of Sistan and Baluchestan, Zahedan, Iran.

*Corresponding Author Email: rvaghee@shahroodut.ac.ir

Received: 16 September 2020, accepted: 17 December 2020

ABSTRACT

Drought intensifies the water crisis and causes irreparable damage to communities. In recent years, machine learning methods have been considered by researchers for drought assessment. The goal of this study is drought assessment in Zabol and Zahedan cities during (1990-2020). In this regard, standardized precipitation index (SPI) at seasonal and annual time scales, partial autocorrelation analysis (PACF), and random forest algorithm (RF) were employed. After SPI calculation, the results of PACF analysis for SPI are used as inputs of the model. Training and testing data were verified with two different inputs. According to the results on a seasonal and annual scale of SPI, the normal drought category (N) has almost the highest frequency in both stations and according to the PACF analysis, the study areas have undergone climate change over 30 years. The results of the model were evaluated by several statistical parameters.

The index of agreement (IOA) of Zabol station for training data by considering four time-lags (2, 4, 6, and 12 months) and three time-lags (2, 4, and 6 months) as input, were 0.9648 and 0.9256, respectively and for testing data was approximately 0.8556 and 0.8673, respectively. IOA at Zahedan station for training data with four time-lags (2, 4, 6, and 8 months) and three time-lags (2, 4, and 6 months) were 0.9495 and 0.9205, respectively, and for testing data were 0.7408 and 0.6303, respectively. Other statistical parameters also indicate the permissible accuracy of the RF model in SPI estimation.

Keywords: Drought, Standardized precipitation index, Random forest algorithm, Partial autocorrelation analysis

HOW TO CITE THIS ARTICLE

Torabi, Z.; Vaghee, R.; Ghaemi, A. (2021). Drought assessment using standardized precipitation index and random forest algorithm. *J. Meteorol. Atmos. Sci.*, 3(4): 390-405

COPYRIGHTS

Copyright for this article is retained by the author(s), with publication rights granted to the JMAs Journal. This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>).

