

Rainfall-runoff Modelling of Coastal Watersheds near Hormuz Strait Using Data Mining

MOJTABA MOHAMMADI¹, HASSAN VAGHARFARD¹, RASOOL MAHDAVI NAJAFABADI^{1*}, PEYMAN DANESHKAR ARASTEH² AND MOHAMMAD JAFAR NAZEMOSADAT³

1. Department of Natural Resources Engineering, Faculty of Agriculture and Natural Resources, University of Hormozgan, Bandar-Abbas, Hormozgan, Iran.

2. Department of Water Sciences and Engineering, Faculty of Agriculture and Natural Resources, Imam Khomeini International University, Qazvin, Iran.

3. Department of Water Engineering, Faculty of Agriculture, Shiraz University, Shiraz, Iran.

(Received: Sep. 8, 2020- Revised: Nov. 3, 2020, Accepted: Jan. 4, 2021)

ABSTRACT

Estimating runoff created by rainfall is a very important step in water resources planning, especially in ungauged River Basins. Therefore, research on models simulating the river flow with minimum error in the river basins is necessary. In this study, rainfall-runoff simulation of Minab watershed was done using data mining methods and their performance was compared to present the proper one. For this purpose, eight data mining algorithms including Model Tree (MT), Random Forest (RF), Support Vector Machines (SVM), Bayesian Ridge Regression (BRR), Gaussian Process (GP), Extreme Gradient Boosting (XGB), Artificial Neural Network (ANN), and Multivariate Adaptive Regression Splines (MARS) were used. Coefficient of determination (R^2), root mean square error (RMSE), mean absolute error (MAE) and Taylor diagram were used to evaluate the model performance. The results indicated that the MARS model had the best performance among the all models to simulate the monthly discharge of the Minab watershed. Also, the SVM model with (RSME =7.73) has a good performance. The other models also performed relatively close to each other (The XGB model with 9.98 had the highest and the MARS model with 7.7 had the lowest RMSE). Then, by entering the values of sea level temperature (PGSST) in the simulation process, the effect of this parameter on the simulation results was investigated. The results showed that PGSST values did not improve the runoff simulation results in the study area.

Keywords: Rainfall-Runoff, Data mining, Minab watershed, Sea Surface Temperature, Persian Gulf.

مدل سازی بارش-رواناب آبخیزهای مناطق ساحلی در نزدیکی تنگه هرمز با استفاده از روش های داده کاوی

مجتبی محمدی^۱، حسن وقار فرد^۱، رسول مهدوی نجف آبادی^{۱*}، پیمان دانش کار آراسته^۲، سید محمد جعفر ناظم السادات^۲

۱. گروه مهندسی منابع طبیعی، دانشکده کشاورزی و منابع طبیعی، دانشگاه هرمزگان، بندرعباس، ایران.

۲. گروه علوم و مهندسی آب، دانشکده کشاورزی و منابع طبیعی، دانشگاه بین المللی امام خمینی (ره)، قزوین، ایران.

۳. گروه مهندسی آب، دانشکده کشاورزی، دانشگاه شیراز، شیراز، ایران.

(تاریخ دریافت: ۱۳۹۹/۶/۱۸ - تاریخ بازنگری: ۱۳۹۹/۸/۱۳ - تاریخ تصویب: ۱۳۹۹/۱۰/۱۵)

چکیده

برآورد رواناب ناشی از وقوع بارندگی، گامی بسیار مهم در برنامه ریزی منابع آب به ویژه در آبخیزهای فاقد ایستگاه های هیدرومتری است. بنابراین پژوهش در ارتباط با مدل هایی که بتواند در این حوضه ها و با کمترین خطا، جریان رودخانه را شبیه سازی نمایند یک ضرورت اجتناب ناپذیر است. در این مطالعه به شبیه سازی بارش-رواناب آبخیز سد استقلال میناب با استفاده از روش های داده کاوی و مقایسه عملکرد آنها و ارائه مناسب ترین مدل بارش-رواناب برای این منطقه پرداخته شد. برای این منظور از هشت مدل داده کاوی شامل الگوریتم جنگل تصادفی، ماشین بردار پشتیبان، شبکه عصبی مصنوعی، مدل الگوریتم های ارتقای شدید گرادیان، مدل درختی MS، مدل اسپلاین چند متغیره رگرسیون انطباقی، مدل فرایند گوسی، مدل بیزی جمعی رگرسیون درختی استفاده گردید. به منظور ارزیابی مدل های مورد استفاده در این تحقیق از معیارهای ارزیابی ضریب تعیین، ریشه میانگین مربعات خطا و میانگین خطای مطلق و همچنین نمودار تیلور استفاده شده است. نتایج نشان داد که مدل اسپلاین چند متغیره رگرسیون انطباقی، بهترین عملکرد را در بین مدل ها برای شبیه سازی دبی ماهانه آبخیز مورد مطالعه داشته است. مدل ماشین بردار پشتیبان نیز با مقدار خطای جذر میانگین مربعات (RMSE) برابر ۷/۷۳ مترمکعب در ثانیه عملکرد مناسبی داشته است. بقیه مدل ها نیز عملکرد نسبتاً نزدیک به هم داشته اند، به طوری که مدل الگوریتم های ارتقای شدید گرادیان با مقدار ۹/۹۸ مترمکعب در ثانیه بالاترین و مدل اسپلاین چند متغیره رگرسیون انطباقی با مقدار ۷/۷ مترمکعب در ثانیه کمترین مقدار RMSE را داشته اند. در ادامه با وارد نمودن مقادیر دمای سطح دریا خلیج فارس (PGSST) به فرایند شبیه سازی به بررسی اثر این پارامتر بر نتایج شبیه سازی پرداخته شد. نتایج نشان داد که مقادیر PGSST موجب بهبود نتایج شبیه سازی رواناب در منطقه مورد مطالعه نگردید.

واژه های کلیدی: بارش-رواناب، داده کاوی، میناب، دمای سطح دریا، خلیج فارس.

مقدمه

آنها بسیار محدود است؛ به طوری که مناسب برای اهداف برنامه ریزی منابع آبی نیستند (Dastorani et al., 2018). لذا، مدیران منابع آب به منظور برنامه ریزی به مدل های مختلف پیش بینی جریان رواناب وابسته هستند (Razavi and Coulibaly, 2013; Luce, 2014). طی دهه های اخیر، بهبود پیش بینی های مدل های ریاضی بارش-رواناب در آبخیزها به ویژه حوضه های فاقد ایستگاه اندازه گیری، هدف اصلی هیدرولوژیست ها بوده است که منجر به پیشرفت های سریع در این زمینه شده است (Hrachowitz et al., 2013). به طور کلی، مدل های ریاضی شناخته شده بارش-رواناب را می توان در دو گروه مدل های دارای مبنای فیزیکی^۱ (همچون مدل های هیدرولوژیکی مفهومی) و مدل های داده محور^۲ (همچون

به طور کلی بخش قابل توجهی از نزولات آسمانی با توجه به ویژگی های فیزیوگرافی آبخیزها به رواناب تبدیل می شود. رواناب حاصل از بارش و ذوب برف یکی از منابع اصلی برای تحقق نیازهای کشاورزی، صنعتی و مصارف خانگی است؛ به طوری که برنامه ریزان منابع آب با تکیه بر داده های جریان رواناب، تخصیص منابع آب به این مصارف را تعیین می کنند (Worland et al., 2018). بنابراین، برآورد رواناب ناشی از وقوع بارندگی، گامی بسیار مهم در برنامه ریزی منابع آب است. تعداد زیادی از آبخیزها در ایران و بیشتر کشورهای در حال توسعه، فاقد ایستگاه های هیدرومتری هستند و یا داده های اندازه گیری شده از رواناب در

* نویسنده مسئول: ra_mahdavi2000@yahoo.com

1 Physically Based Models
2 Data-Driven Models

نتایج پیش‌بینی جریان روزانه و ماهانه دبی در حوضه‌های مورد مطالعه نسبت به مدل‌های هیدرولوژیکی جعبه خاکستری شده است. (Samadi et al., 2019) نیز به منظور پیش‌بینی دبی ماهانه ورودی به سد بوستان در استان گلستان از مدل‌های داده‌کاوی و ترکیبی استفاده نمودند. برای این منظور از مدل‌های شبکه عصبی مصنوعی، ماشین بردار پشتیبان و سری‌های زمانی استفاده نمودند. یافته‌های این محققین نشان داد که مدل سری زمانی عملکرد مناسب‌تری را نسبت به بقیه مدل‌ها داشته است. (Ahmadi, 2019) به بررسی ارزیابی عملکرد روشی‌های ماشین‌بردار پشتیبان (SVM) و سیستم استنتاج عصبی فازی تطبیقی (ANFIS) در پیش‌بینی جریان ماهانه پرداخت. نتایج نشان داد که روش SVM از عملکرد بهتری نسبت به مدل ANFIS در پیش‌بینی جریان برخوردار است.

طی بررسی‌های به‌عمل‌آمده، تغییرات در حجم منابع آب نیز در ارتباط با تغییرات اقلیمی ناشی از نوسانات گرما و رطوبت نیوار در الگوهای پیوند از دور است؛ (Bergström et al., 2001; Evans et al., 1998). مفهوم پیوند از دور ابتدا به‌وسیله Ångström (1935) ارائه گردید که ارتباط بین تغییرات جریان‌های نیواری و ناهنجاری‌ها را از راه دور توصیف می‌نماید (Wang et al., 2019). به بیانی دیگر گاهی اجزای سیستم اقلیمی اثرات خود را بر مکانی دورتر از محل وقوع آن تحمیل می‌کنند که این گونه تأثیرگذاری به "پیوند از دور" موسوم است. پیوند از دور می‌تواند سازوکار فیزیکی و علت پدیده‌های جریان رواناب را روشن سازد. در دهه اخیر، استفاده از فنون پیوند از دور به‌صورت فزاینده‌ای در زمینه‌های مختلف هیدرولوژی همچون پیش‌بینی رواناب (Peters et al., 2013)، محاسبه حجم توده برف (Sobolowski and Frei, 2007)، آنالیز و پیش‌بینی بارش (He and Guan, 2013; Kirono et al., 2010; Mekanik et al., 2015)، تخمین میزان دی‌اکسید کربن نیوار (Chiew et al., 1998) و نظایر آنها، رواج یافته‌اند. نوسان‌های دمای سطح دریا (SST) یکی از جالب توجه‌ترین نمونه‌های سازوکار پیوند از دور است. این نمایه یکی از نمایه‌های ارزشمند و پایه‌ای در ارزیابی پیوند اقلیمی دریا و خشکی است (Heburn, 1985) که می‌تواند به پیش‌بینی رواناب کمک کند (Sundy et al., 2014)، به‌طوری‌که در سال‌های اخیر، مطالعات متعددی بر روی بررسی شرایط اولیه و مرزی درجه حرارت سطح دریا که قادر به تأثیرگذاری بر روی رواناب مناطق ساحلی هستند، متمرکز شده‌اند (Jansa et al., 2014; Drobinski et al., 2018)

شبکه عصبی مصنوعی و ماشین بردار^۱ تقسیم‌بندی نمود (Farmer and Vogel, 2016; Niu et al., 2019). مدل‌های توزیعی دارای مبنای فیزیکی که به‌عنوان مدل‌های جعبه سفید^۲ (Yaseen et al., 2016) نیز شناخته می‌شوند، فرآیندهای هیدرولوژیکی را با دقت بالا بیان می‌کنند و به‌عنوان یک استاندارد برای مدل‌سازی بارش-رواناب در نظر گرفته می‌شوند. مدل‌های داده‌محور که مدل‌های جعبه سیاه^۳ (Yaseen et al., 2016) نیز نامیده می‌شوند، برای مناطق و حوضه‌هایی که داده‌های اندازه‌گیری شده از محیط فیزیکی آنها دارای محدودیت است، اما داده‌های بارش و رواناب با طول دوره آماری بالا دارند، یک ابزار مناسب نیستند (Iorgulescu and Beven, 2004). توسعه فنون پیچیده یادگیری ماشین^۴ همراه با افزایش سریع توانایی‌های محاسباتی رایانه‌ای منجر به گسترش پژوهش‌ها در زمینه روش‌های پیشرفته پیش‌بینی رواناب مبتنی بر مدل‌های داده‌محور در دهه‌های گذشته شده است (Shortridge et al., 2016). بر اساس تقسیم‌بندی Yaseen و همکاران (۲۰۱۵)، روش‌های پیش‌بینی جریان مبتنی بر مدل‌های داده‌محور در چهار گروه یادگیری ماشین، مجموعه‌های فازی، محاسبات تکاملی و مدل‌های ترکیبی موجک تقسیم‌بندی می‌شوند. در سال‌های اخیر، روش‌های داده‌کاوی به علت آن‌که توانایی مدل‌سازی فرایندهای غیرخطی و غیر ایستای جریان رودخانه را بدون نیاز به اطلاعات فیزیکی آن دارند، در هیدرولوژی و مدیریت منابع آب مورد توجه پژوهشگران قرار گرفته است. به‌عنوان نمونه از فنون داده‌کاوی در شبیه‌سازی بارش-رواناب (Worland et al., 2018; Dastorani et al., 2018; Niu et al., 2019; Sezen et al., 2019; Feng et al., 2019)؛ پیش‌بینی سیلاب (Eini et al., 2020)؛ جریان رسوب (Kumar et al., 2019; Khosravi et al., 2020)؛ فرسایش آبی (Pourghasemi et al., 2017)؛ (Pourghasemi and Rahmati, 2018) تبخیر و تعرق (Granata, 2020)؛ منشأیابی گرد و غبار (Gholami et al., 2020) و دمای سطح دریا (Wolf et al., 2020) ارائه شده است. در ایران نیز (Sattari et al., 2016) به شبیه‌سازی دبی سیلاب رودخانه اهرچای در مقیاس ساعتی با استفاده از روش یادگیری ماشین پرداختند، یافته‌ها نشان داد که روش رگرسیون ماشین بردار پشتیبان بهترین دقت را در بین مدل‌ها دارد. یافته‌های (Ghorbani et al., 2016) و (Bagherpour et al., 2019) نیز نشان داد که استفاده از روش‌های داده‌محور منجر به افزایش قابل‌ملاحظه‌ای در دقت

و مقایسه عملکرد آنها و ارائه مناسب‌ترین مدل بارش-رواناب برای این منطقه پرداخته می‌شود. در ادامه با وارد نمودن متغیر SST به عنوان یک ورودی در کنار ورودی‌های دیگر مدل به بررسی این پرسش پرداخته می‌شود که ورود متغیر SST به فرایند مدل‌سازی بارش-رواناب، چگونه موجب بهبود نتایج مدل‌سازی می‌شود؟ به بیان دیگر با توجه به در اختیار بودن داده‌های به‌روز SST آیا می‌توان از این متغیر به‌عنوان متغیر اصلی در حوضه‌های فاقد اطلاعات هیدرومتری به منظور شبیه‌سازی بارش-رواناب استفاده نمود یا خیر؟

مواد و روش‌ها

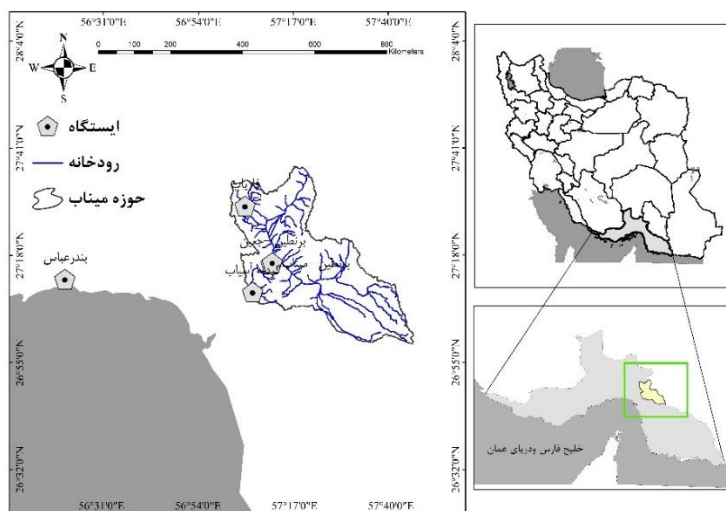
منطقه مورد مطالعه

استان هرمزگان با مساحتی حدود ۷۲ هزار کیلومترمربع در شمال تنگه هرمز واقع شده است. میانگین نزولات آسمانی سالانه این استان کمتر از ۱۸۵ میلی‌متر با پراکنش زمانی و مکانی ناهمگون و غالباً به صورت رگبارهای شدید به همراه سیلاب‌های مخرب و زودگذر است (Bayat varkashi and Gheysari, 2018). رودخانه میناب در ۹۵ کیلومتری شرق بندرعباس با مساحت آبخیز ۱۰/۲۰۰ کیلومترمربع (Nohegar et al, 2016) و در فاصله بین ۴۸' ۵۶° تا ۵۹' ۵۹° شرقی و ۲۷° تا ۳۲' ۲۸° عرض شمالی پس از به هم پیوستن رودخانه‌های جگین و رودان در فاصله حدود ۲۰ کیلومتری سد استقلال به‌وجود می‌آید (Nohegar et al, 2013). این آبخیز از بزرگ‌ترین و مهم‌ترین آبخیزهای سواحل دریای عمان و خلیج فارس در استان هرمزگان است که از لحاظ منابع آب و کشاورزی یکی از حوضه‌های پرآب محسوب شده و آب آن نیز از لحاظ شرب دارای کیفیت مناسبی است. شکل (۱)، موقعیت آبخیز میناب را نشان می‌دهد.

توپوگرافی پیچیده بوده (Senator et al., 2020) و بارش‌های فصلی منشأ اصلی رواناب این مناطق هستند (Rasoli et al., 2012)، قوی‌تر است. در ایران نیز (Nazemosadat et al (1995) و Nazemosadat (1998) روشن ساختند که نوسان‌های دمای سطح آب خلیج فارس نشان معنی‌داری بر نوسان‌های بارش در بسیاری از گستره‌های جنوب غربی و جنوبی ایران دارد. آب‌های خلیج فارس و دریای عمان افزون بر آن که نقش ارزشمندی در زندگی اقتصادی-اجتماعی مردم ایران دارند، پیشینه پرننگی نیز در پیدایش الگوهای گوناگون آب و هوایی در ایران و به‌ویژه پهنه های جنوبی دارند (Nazemosadat et al, 2014). بررسی تحقیقات صورت‌گرفته بیانگر آن هستند که شناخت نوسان‌های دمای سطح آب در پهنه‌های آبی خلیج فارس می‌تواند در مدیریت منابع آب و فعالیت‌های زیست‌محیطی در نواحی جنوبی کشور به‌ویژه سواحل حاشیه خلیج فارس و تنگه هرمز استفاده شوند.

میانگین بارندگی مناطق جنوبی کشور در نزدیکی سواحل تنگه هرمز کمتر از میانگین کشوری است. به‌طوری‌که این مقدار نیز در سال‌های اخیر به دلیل تغییرات اقلیمی کاهش یافته است، به‌طوری‌که مخازن و سدهای آبی موجود در منطقه پاسخ‌گوی نیاز فعلی منطقه نیست. در این بین، آب‌های زیرزمینی دشت میناب و سد استقلال به عنوان تأمین‌کننده ۵۰ درصد آب موردنیاز بندرعباس و اراضی کشاورزی میناب از اهمیت ویژه برخوردار است. از سوی دیگر رشد و توسعه زیاد شهرستان بندرعباس در سال‌های اخیر موجب شده است که نیاز منابع آبی در این مناطق روز به روز بیشتر شود. لذا، تعیین آب ورودی و پیش‌بینی مقدار آن در رودخانه‌های این مناطق یکی از چالش‌های پیش روی این مناطق است. بنابراین در این مطالعه در ابتدا با هدف پیش‌بینی مقدار آب ورودی به سد و تأمین آب آبیاری به شبیه‌سازی بارش

رواناب آبخیز سد استقلال میناب با استفاده از روش‌های داده‌کاوی



شکل ۱- موقعیت حوضه آبخیز مورد مطالعه

داده‌ها

داده‌های هیدرومتئولوژیکی

ماهانة ایستگاه‌های دارای دوره داده‌برداری ۱۳۴۶-۱۳۹۴ و همچنین به‌منظور نتایج بهتر از داده‌های ایستگاه سینوپتیک بندرعباس نیز استفاده گردید. جدول (۱)، مشخصات ایستگاه‌ها و پارامترهای ورودی مورد استفاده را نشان می‌دهد.

در این مطالعه به‌منظور شبیه‌سازی شدت جریان ایستگاه هیدرومتری برنطین، از داده‌های هیدرومتری و تبخیرسنجی

جدول ۱- موقعیت ایستگاه‌ها و پارامترهای ورودی مورد استفاده در تحقیق

نام ایستگاه	نوع ایستگاه	ارتفاع از سطح دریا (متر)	طول جغرافیایی	عرض جغرافیایی	پارامتر	مقدار متوسط سالانه
برنطین	هیدرومتری	۱۳۰	۵۷° ۱۱' ۳۱"	۲۷° ۱۶' ۱۷"	دبی (m ³ /s)	۸/۱۴
برنطین	تبخیرسنجی	۱۴۰	۵۷° ۱۱' ۴۷"	۲۷° ۱۶' ۳۸"	تبخیر (mm)	۲۸۵
فاریاب	تبخیرسنجی	۳۰۰	۵۷° ۰۵' ۱۵"	۲۷° ۲۸' ۴۷"	بارش (mm)	۲۳۵
					دما (°C)	۲۸/۸
					بارش (mm)	۲۷۲
					دمای متوسط (°C)	۲۶/۹
					دمای حداکثر (°C)	۳۲/۲
					دمای حداقل (°C)	۲۱/۶
بندرعباس	سینوپتیک	۹/۸	۵۶° ۲۲' ۰۰"	۲۷° ۱۳' ۰۰"	رطوبت متوسط (%)	۶۴/۸
					رطوبت حداکثر (%)	۸۳/۶
					رطوبت حداقل (%)	۴۵/۱

داده‌های دمای سطح آب Sea Surface Temperature

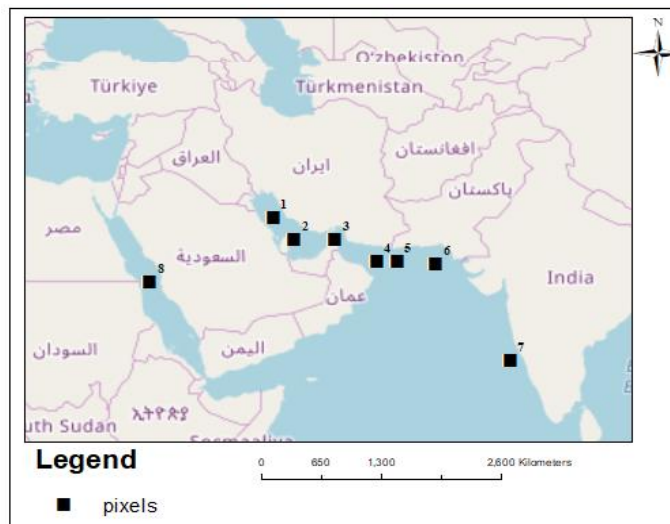
در نزدیکی تنگه هرمز پرداختند و برای این منظور سه سلول SST در خلیج فارس، دو سلول در دریای عمان، دو سلول در دریای عرب و یک سلول در دریای سرخ را مورد بررسی قرار دادند و مؤثرترین سلول را بر بارش ایستگاه سینوپتیک بندرعباس در ماه‌ها و فصول مختلف ارائه نمودند. جدول (۲)، نتایج پژوهش این محققین و شکل (۲)، موقعیت هر کدام از سلول‌ها را نشان می‌دهد.

در این مطالعه، علاوه بر داده‌های هیدرومتئولوژیکی به‌عنوان ورودی مدل از داده‌های دمای سطح آب خلیج فارس نیز استفاده شد. برای این منظور از نتایج پژوهش Mohammadi, (2020) استفاده گردید. این محققین به بررسی تأثیر نوسان‌های دمای سطح آب خلیج فارس بر نوسان‌های بارش در پهنه ساحلی ایران

جدول ۲- مؤثرترین سلول SST بر بارش ایستگاه سینوپتیک بندرعباس (منبع: Mohammadi, 2020)

ماه‌های سال	JAN	FEB	MAR	APR	OCT	NOV	DEC	JFM	FMA	OND
مؤثرترین سلول	سلول ۱ و ۳	سلول‌های ۳ و ۸	سلول ۱ با ۲	سلول ۳	سلول ۷ با ۵	سلول ۴ با ۳	سلول ۸ با ۵	سلول ۱ با ۲	سلول ۱ با ۲	سلول ۵ با ۳

OND اکتبر، نوامبر، دسامبر؛ JFM ژانویه، فوریه، مارس؛ FMA فوریه، مارس، آوریل؛ همچنین منظور از شیب SST پیکسل‌ها، شیب اختلاف دمایی بین دو سلول SST است.



شکل ۲- موقعیت سلول‌های SST (منبع: Mohammadi, 2020)

مدل‌های مورد استفاده

2000 ارائه شده است. در مطالعه حاضر از بسته LiblineR (Helleputte and Gramme, 2017) نرم‌افزار آماری R استفاده گردید.

شبکه عصبی مصنوعی

شبکه‌های عصبی مصنوعی یک مدل از روش‌های یادگیری ماشینی است که الهام گرفته از شبکه‌های عصبی بدن انسان است (Aleotti and Chowdhury, 1999). شبکه‌های عصبی مصنوعی شامل طیف وسیعی از الگوریتم‌های مختلف بوده که در هوش مصنوعی و آمار توسعه یافته است (Pourghasemi and Rahmati, 2018) و از آنجایی که از یک گروه به هم پیوسته از نوروهای مصنوعی تشکیل شده‌اند، توانایی یادگیری روابط پیچیده بین متغیرهای ورودی و خروجی را داشته (Fausett, 1994) و قادر به تجزیه و تحلیل و پیش‌بینی پدیده‌هایی را که ماهیت غیرخطی دارند، دارا می‌باشند (Peddle et al., 1994). جزئیات کاربرد روش شبکه عصبی مصنوعی در مدل‌سازی بارش- رواناب در مطالعات مختلف از جمله (Zema et al., 2020) و (Riad et al., 2004) بیان شده است. در مطالعه حاضر از بسته brnn (Rodriguez and Gianola, 2020) نرم‌افزار آماری R استفاده گردید.

مدل الگوریتم‌های ارتقای شدید گرادیان^۷

این مدل یک رویکرد یادگیری ماشینی گروهی است که از مجموعه یادگیرهای ضعیف^۸، یک یادگیرنده قوی^۹ می‌سازد (Chemura et al., 2020). مدل XGBoost یک گام بهبود یافته از روش جزءبندی مبتنی بر درخت بازگشتی^{۱۱} ارتقای گرادیان است که توسط Friedman (2001) معرفی گردید. پس از توسعه توسط Chen and Guestrin (2016) به‌عنوان یکی از الگوریتم‌های بسیار کارآمد یادگیری ماشینی برای زمینه‌های مختلف شناخته شده (Chemura et al., 2020) و در مدل‌های رگرسیونی و طبقه‌بندی مورد استفاده گسترده قرار گرفته‌است (Nobre and Neves, 2019; Gholami et al., 2020). برای دریافت جزئیات به مطالعه Friedman (2001) و Chen and Guestrin (2016) و Nobre and Neves (2019) مراجعه گردد. در مطالعه حاضر از بسته xgboost (Chen et al., 2020 (a)) نرم‌افزار آماری R استفاده گردید.

مدل درختی M5^{۱۲}الگوریتم جنگل تصادفی^۱

الگوریتم جنگل تصادفی ابتدا با جایگزینی و تغییر مداوم عوامل مؤثر و مرتبط با هدف، تعداد زیادی درخت تصمیم‌گیری شده ایجاد می‌نماید و سپس ترکیب این درختان فردی با هم موجب کاهش واریانس و افزایش دقت پیش‌بینی‌ها می‌گردد (Breiman, 2001). در این روش، هر درخت تصمیم بر روی نمونه بوت‌استرپ^۲ که با نمونه‌برداری همراه با جایگزینی داده اولیه به‌دست آمده است، توسعه داده می‌شود. برخلاف بیشتر روش‌های مبتنی بر یادگیری ماشین، جنگل تصادفی فقط به دو پارامتر برای تهیه یک مدل پیش‌بینی نیاز دارد (Rodriguez-Galiano et al., 2015). تعداد درختان رگرسیون (K) و عوامل (x) (Gholami et al., 2020) که در هر گره به‌منظور رشد درختان رگرسیون، مورد استفاده قرار می‌گیرند (Rodriguez-Galiano et al., 2012). در چند کلمه می‌توان بیان داشت که جنگل تصادفی شامل دو ایده قدرتمند شامل انتخاب ویژگی تصادفی و کیسه بندی^۳ در الگوریتم‌های یادگیری ماشینی است (Wu et al., 2014). برای اطلاعات بیشتر به مطالعات Breiman (2001) و Liaw و Wiener (2002) می‌توان مراجعه نمود. در این مطالعه از پکیج party نرم‌افزار آماری R استفاده شد.

ماشین بردار پشتیبان

ساختار مدل ماشین پشتیبان یک طبقه‌بندی نظارت شده است که از اصل استقرایی کمینه‌سازی خطر ساختاری^۴ استفاده نموده و برای دو گروه طبقه‌بندی و رگرسیون می‌تواند استفاده شود (Vapnik, 2000). مدل طبقه‌بندی ماشین بردار پشتیبان برای حل مسائل طبقه‌بندی داده‌هایی که در کلاس‌های مختلف قرار می‌گیرند و از مدل رگرسیون ماشین بردار پشتیبان در حل مسائل پیش‌بینی استفاده می‌شود (Shawe-Taylor and Cristianini, 2000). دقت پیش‌بینی یک مدل SVM متأثر از انتخاب نوع تابع کرنال همچون حلقوی^۵، چندجمله‌ای^۶، خطی و شعاعی (RBF)^۷ است (Pourghasemi and Rahmati, 2018). تابع کرنال RBF که بر پایه فاصله اقلیدسی استوار است، بیشترین کاربرد را در مدل‌سازی‌های زیست‌محیطی دارد (Marjanović et al., 2011; Tien Bui et al., 2012). توضیح آماری بیشتر مدل SVM در مطالعات Vapnik (1998); Shawe-Taylor and Cristianini,

7 Radial Basis Function

8 Extreme Gradient Boosting (XGBoost)

9 Weak Learners

10 Strong Learner

11 The Recursive Tree-Based Partitioning Method

12 M5 Tree Model

1 Random Forest (RF)

2 Bootstrap

3 Bagging

4 Structural Risk Minimization

5 Sigmoid

6 Polynomial

(Yang et al., 2003). یک تابع پایه در یک بازه معین تعریف شده و نقاط ابتدایی و انتهایی آن گره نامیده می شود. دو گام اصلی در توسعه مدل MARS به منظور ایجاد یک ساختار بهینه برای پیش بینی پدیده ها وجود دارد. در گام اول همه توابع هدف ممکن تعریف می گردد. در نتیجه، مدلی بسیار پیچیده بیش برآزش شده^۵ با تعداد گره های زیاد به دست می آید. در گام دوم برای جلوگیری از بیش برآزش از طریق فرایند هرس کردن و با استفاده از روش GCV توابع پایه که اهمیت کمتری دارند، شناسایی و حذف می شوند (Samadi et al., 2015). برای دریافت جزئیات به مطالعه Friedman (1991) و Zahiri and Nezaratian (2020) مراجعه گردد. در مطالعه حاضر از بسته earth (Milborrow, 2020) نرم افزار آماری R استفاده گردید.

مدل فرایند گاوسی

فرایند گاوسی^۶ بر اساس فرایندهای تصادفی^۷ استوار است که یک ابزار قدرتمند برای استنتاج احتمالاتی مستقیم^۸ توزیع ها بر توابع است (O'Hagan, 1978). در فرایند گاوسی، $f(x)$ تعمیمی طبیعی از توزیع گاوسی است و به طور کامل توسط تابع میانگین $m(x)$ و تابع کوواریانس آن معین می شود (Li et al., 2017) که به صورت زیر تعریف می شود:

$$m(x) = E(f(x)) \quad (\text{رابطه ۲})$$

$$k(x, x') = E((f(x) - m(x))(f(x') - m(x')))) \quad (\text{رابطه ۳})$$

که در آن، $k(x, x')$ کوواریانس (یا تابع کرنال) که در x و x'

ارزیابی می گردد و $f(x)$ به صورت زیر ارائه می گردد:

$$f(x) \sim GP(m(x), k(x, x')) \quad (\text{رابطه ۴})$$

در فرایند گاوسی هر مشاهده y را می توان از طریق مدل نویزی گاوسی به یک تابع اصلی ارتباط داد:

$$y_i = f(x_i) + \varepsilon \quad (\text{رابطه ۵})$$

که در آن، $\varepsilon \sim N(0, \sigma_f^2)$ به عنوان نویز تابع توزیع نرمال با میانگین صفر و واریانس σ_f^2 است. رگرسیون در واقع جستجو برای $f(x)$ است. برای سادگی کار در مرحله بعدی با نوشتن عبارت زیر از روش جدید آمیختن نویز در $K(x, x')$ استفاده می شود:

$$k(x, x') = \sigma_f^2 \exp\left[-\frac{(x-x')^2}{2\tau^2}\right] + \sigma_n^2 \delta(x, x') \quad (\text{رابطه ۶})$$

که در آن، $\delta(x, x')$ تابع دلتای کرنر است. اطلاعات

بیشتر در خصوص فرایند گاوسی در پژوهش های Rasmussen and Williams, 2006; Seeger, 2004; Li et al., 2017

مدل درختی M5 که ابتدا به وسیله (Quinlan 1992) معرفی گردید، زیرمجموعه ای از مدل های بر پایه درختی است که سعی بر پیش بینی متغیرهای عددی با استفاده از مدل های رگرسیونی دارد (Nourani et al., 2019); به طوری که این الگوریتم، درخت تصمیم را با رگرسیون خطی ترکیب نموده و از مدل رگرسیون چند متغیره برای برگ های درختان رگرسیون استفاده می نماید (Wang and Witten, 1996; Etemad-Shahidi and Bonakdar, 2009). مدل درخت M5 در ابتدا یک درخت می سازد و سپس اطلاعات را از آن استخراج می نماید (Etemad-Shahidi and Taghipour, 2012). به این صورت که این الگوریتم در ابتدا با استفاده از داده های ورودی یک درخت معکوس می سازد و رگرسیون خطی را برای هر برگ ایجاد می کند که مناسب برای یک دامنه خاص از فضای ورودی است (Bhattacharya and Solomatine, 2005). ساختن، هرس نمودن و هموارسازی سه گام اصلی الگوریتم M5 است (Zahiri and Nezaratian, 2020). این الگوریتم فضای مسئله^۱ را با استفاده از عامل کاهش انحراف معیار^۲ تقسیم بندی می کند.

$$SDR = sd(T) - \sum \frac{T_i}{|T|} \times sd(T_i) \quad (\text{رابطه ۱})$$

که در آن، T مجموعه ای از داده های ورودی است که به گره^۳ ختم می شوند، T_i مجموعه هایی که از تقسیم گره با توجه به ویژگی انتخاب شده حاصل می شوند و sd انحراف معیار است (Wang and Witten, 1996; Zahiri and Nezaratian, 2020). در مطالعه حاضر از بسته RWeka (Hornik et al., 2020) نرم افزار آماری R استفاده گردید.

مدل اسپلاین چند متغیره رگرسیون انطباقی^۴

مدل اسپلاین چند متغیره رگرسیون انطباقی (MARS) توسط Friedman (1991) به عنوان یک الگوریتم انعطاف پذیر برای سازماندهی ارتباط بین پارامترهای ورودی و متغیر هدف و پیش بینی آن ارائه گردید (Zhang and Goh, 2016). مدل MARS یک روش با دقت بالا برای یافتن ارتباط بین متغیرهای ورودی و خروجی از یک مجموعه داده است (Haghiabi, 2016). این الگوریتم به دنبال یک رابطه غیر خطی در یک مجموعه بزرگ داده ها است تا با توجه به هدف مربوطه، وزن متغیرهای ورودی را به منظور ایجاد یک مدل واضح، تعیین نماید و توابع پایه ای را برای پارامترهای ورودی ایجاد می کند و سپس با ترکیب توابع پایه برای تخمین پارامتر خروجی، مدل های رگرسیون را استخراج می کند

6 Gaussian Process
7 Stochastic Process¹
8 Probabilistic Inference Directly

1 Problem Space
2 Standard Deviation Reduction (Sdr)
3 Node
4 Multivariate Adaptive Regression Splines
5 Overfitted Models

استفاده است. در مطالعه حاضر از بسته (Karatzoglou) Kernlab (et al., 2019) نرم افزار آماری R استفاده گردید.

مدل بیزی جمعی رگرسیون درختی

مدل بیزی جمعی رگرسیونی^۱ درختی اولین بار در اجلاس با موضوع پیشروی و پیشرفت های تداخل در سیستم های پردازش اطلاعات عصبی در سال ۲۰۰۶ بیان شد. این مدل یک روش رگرسیون ناپارامتری بیزی است که مبتنی بر درخت بوده و از روش های گردآوری به صورت کلی و الگوریتم های کمکی به طور ویژه استفاده می کند و با یک مدل آماری تعریف می شود. در این روش بیزی از جمع درخت ها برای مدل کردن یا تقریب زدن $f(x) = E(Y|X)$ استفاده می شود (Hosseini et al, 2019). مدل BART را می توان به صورت رابطه ۷ بیان داشت (Kapelner and Bleich, 2013). (رابطه ۷)

$$Y = f(X) + \varepsilon \approx \tau_1^M(X) + \tau_2^M(X) + \dots + \tau_m^M(X) + \varepsilon, \quad \varepsilon \sim N_n(0, \sigma^2 I_n)$$

که در آن، Y یک بردار n در ۱ از پاسخ ها، X ماتریس طراحی^۲ n * p، ε بردار نویز α در ۱ و M تعداد درختان رگرسیون مجزا است که هر کدام از یک ساختار درختی تشکیل شده اند که با τ نشان داده می شود و پارامترهای موجود در انتهای گره ها که برگ نامیده می شود به وسیله M بیان می شود. این دو با یکدیگر به وسیله τ^M بیان می شوند که نشان دهنده یک درخت کامل با ساختار و مجموعه ای از پارامترهای برگ است. اطلاعات بیشتر در مطالعات (Kapelner and Bleich, 2013) و (Hosseini et al, 2019) بیان شده است. در مطالعه حاضر از بسته monomvn (Gramacy, 2019) نرم افزار آماری R استفاده گردید.

معیارهای ارزیابی

به منظور ارزیابی مدل های مورد استفاده در این تحقیق از معیارهای ارزیابی ضریب تعیین، ریشه میانگین مربعات خطا و میانگین خطای مطلق به شرح زیر استفاده گردید.

$$R^2 = \left[\frac{\sum_{i=1}^n (Q_o - \bar{Q}_o)(Q_s - \bar{Q}_s)}{\sqrt{\sum_{i=1}^n (Q_o - \bar{Q}_o)^2 \sum_{i=1}^n (Q_s - \bar{Q}_s)^2}} \right]^2 \quad (\text{رابطه ۸})$$

$$RMSE = \frac{\sqrt{\sum_{i=1}^n (Q_o - Q_s)^2}}{n} \quad (\text{رابطه ۹})$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |Q_s - Q_o| \quad (\text{رابطه ۱۰})$$

که در آنها، Q_o مقدار دبی مشاهده ای، Q_s دبی شبیه سازی شده، n تعداد داده ها، \bar{Q}_o میانگین دبی های مشاهده ای، \bar{Q}_s میانگین دبی های محاسباتی است.

در این مطالعه، علاوه بر روش های ارزیابی فوق الذکر از نمودار تیلور (Taylor, 2001) نیز به منظور ارزیابی کارایی مدل ها استفاده شد. این نمودار بر مبنای ارتباط هندسی بین ضریب همبستگی، انحراف معیار سری های زمانی و $RMSD^2$ استوار است. شکل (۳)، نمونه نمودار تیلور را نشان می دهد. روابط و اطلاعات بیشتر در مطالعه Taylor (2001) درج گردیده است.

نمودار تیلور به دو صورت نیم دایره (نمایش همبستگی منفی و مثبت) و ربع دایره (فقط نمایش همبستگی مثبت) ارائه می شود که در هر دو صورت، مقادیر ضریب همبستگی به صورت شعاع دایره روی قوس آن، مقادیر انحراف معیار به صورت دایره متحدالمرکز نسبت به مرکز دایره و مقادیر $RMSD$ به صورت دایره متحدالمرکز نسبت به نقطه مرجع (دایره توخالی روی محور افقی) ترسیم می شوند. نقطه مرجع موقعیت داده مشاهده ای را بر اساس انحراف معیار سری زمانی آن نشان می دهد (زیرا مقدار $RMSD$ و ضریب تعیین سری زمانی دبی مشاهده ای در مقایسه با خودش به ترتیب صفر و یک خواهد بود. لذا، موقعیت آن روی محور افقی بر اساس مقدار انحراف معیار تعیین خواهد شد. روش ارزیابی در این نمودار به این صورت است که موقعیت داده های مورد بررسی بر اساس مقدار $RMSD$ ، ضریب همبستگی آن با داده های مشاهده ای و انحراف معیار سری زمانی روی نمودار ترسیم می شود و هر داده ای که موقعیت مکانی آن روی نمودار به نقطه مرجع نزدیک تر باشد، با دقت بیشتری مقادیر سری زمانی را شبیه سازی می نماید.

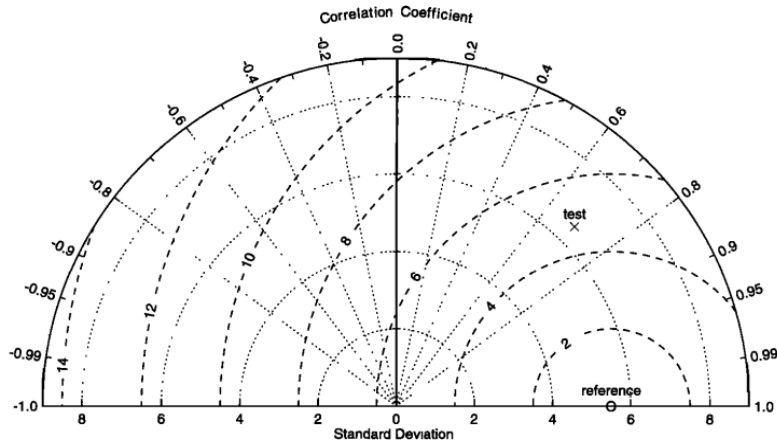
نتایج و بحث

نتایج شبیه سازی دبی ماهانه

در این مطالعه، همان طور که اشاره شد از هشت مدل بر پایه روش های داده کاوی به منظور شبیه سازی دبی ماهانه ایستگاه هیدرومتری آبخیز سد استقلال میناب استفاده گردید. برای این منظور از ۷۰ درصد داده ها برای مرحله آموزش و ۳۰ درصد داده ها برای مرحله صحت سنجی استفاده شده است. جدول (۳)، نتایج ارزیابی مدل های مورد استفاده را در دو مرحله آموزش و صحت سنجی نشان می دهد. چنان که جدول (۳) نشان می دهد، نتایج عملکرد مدل ها در مرحله صحت سنجی بر اساس معیار ارزیابی R^2 به ترتیب با مقادیر ۰/۶۸، ۰/۶۸، ۰/۶۷ مربوط به مدل های MARS، RF و ANN می باشد و بر اساس معیار ارزیابی RMSE بهترین عملکرد به ترتیب با مقادیر ۷/۷ و ۷/۷۳ مترمکعب در ثانیه مربوط به مدل های MARS و SVM است و بر اساس معیار

نمودار تیلور می توان بیان داشت که مدل MARS بهترین عملکرد را در بین مدل ها برای شبیه سازی دبی ماهانه آبخیز مورد مطالعه داشته است. شکل (۵) نیز، نمودار پراکنش تعدادی از مدل ها را نشان می دهد. همان طور که از شکل (۵) مشخص است مدل های کوچکتر را بهتر شبیه سازی نموده اند به طوری که در مقادیر کمتر از ده مترمکعب بر ثانیه پراکنش داده ها حول محور یک به یک بیشتر است و در دبی های بین ۲۰ تا ۵۰ مترمکعب بر ثانیه پراکنش داده های شبیه سازی شده و مشاهداتی از محور یک به یک فاصله بیشتری دارند.

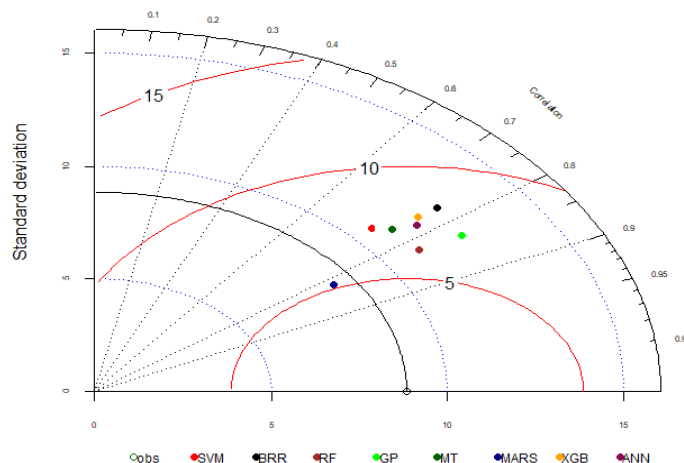
ارزیابی MAE بهترین عملکرد به ترتیب با مقادیر ۵ و ۵/۱۶ مترمکعب در ثانیه مربوط به مدل های SVM و ANN است. همان طور که در بخش قبل اشاره شد در این مطالعه از نمودار تیلور نیز به منظور ارزیابی عملکرد مدل ها استفاده گردید. شکل (۴)، نتایج نمودار تیلور را نشان می دهد. چنان که در این نمودار مشخص است با توجه به این که مدل MARS از نظر موقعیت مکانی نزدیک ترین مدل به نقطه مرجع و نزدیک ترین مدل به دایره RMSD با مقدار ۵ بر روی نمودار است، بهترین عملکرد را دارد. بنابراین با توجه به نتایج ارزیابی جدول (۳) و همچنین نتایج



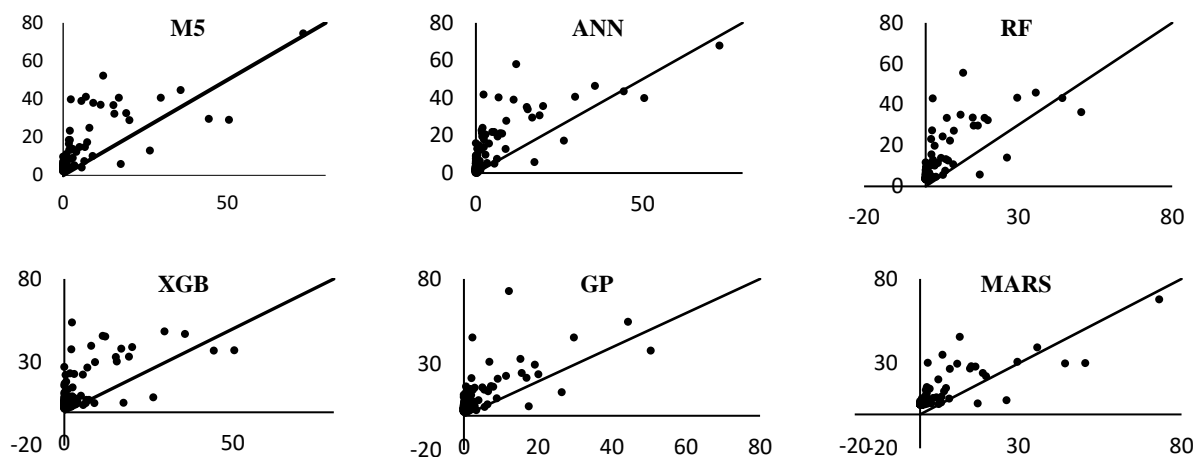
شکل ۳- نمودار تیلور، ضریب همبستگی به صورت شعاع دایره، انحراف معیار و RSMD به ترتیب به صورت دوایر متحدالمرکز نسبت به نقطه مرجع نشان داده شده است. نقطه مرجع، موقعیت ایستگاه زمینی را بر اساس انحراف معیار سری زمانی نشان می دهد (منبع: Taylor, 2001).

جدول ۳- نتایج ارزیابی مدل های مورد استفاده در دو مرحله آموزش و صحت سنجی

صحت سنجی			آموزش			روش	مخفف مدل	مدل
MAE	R2	RMSE	MAE	R2	RMSE			
۵/۸۸	۰/۵۸	۸/۸۶	۶/۹۷	۰/۵۶	۲۰/۷۶	'M5'	MT	مدل درختی
۰/۱۳	۰/۶۸	۸/۵۷	۵/۴۵	۰/۶۵	۱۴/۱۰	'cforest'	RF	الگوریتم جنگل تصادفی
۵	۰/۵۶	۷/۷۳	۵/۵۷	۰/۶۲	۱۶/۱۶	"svmLinear3	SVM	ماشین بردار پشتیبان
۵/۵	۰/۵۷	۹/۸	۷/۰۳	۰/۶۴	۱۸/۵۸	bridge	BRR	بیزی جمعی رگرسیون درختی
۵/۷۹	۰/۷	۸/۵	۶/۱۳	۰/۵۲	۱۶/۶	gaussprRadial	GP	فرایند گوسی
۶/۹۸	۰/۵۲	۹/۹۸	۷/۴۲	۰/۴۳	۲۲/۹۸	xgbLinear	XGB	الگوریتم های ارتقای شدید گرادیان
۵/۱۶	۰/۶۷	۸/۷۶	۷/۰۸	۰/۵۴	۲۰/۱۵	brnn	ANN	شبکه عصبی مصنوعی
۶/۶۵	۰/۶۸	۷/۷	۰/۹۷	۰/۶۱	۲۰/۴۱	bagEarth	MARS	مدل اسپلاین چند متغیره رگرسیون انطباقی



شکل ۴- نتایج نمودار تیلور برای مدل های مورد استفاده



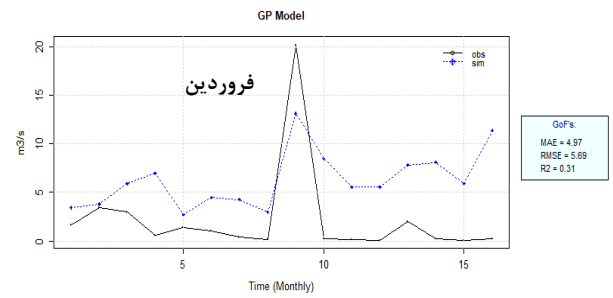
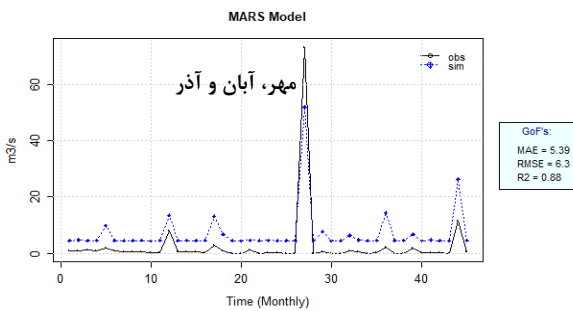
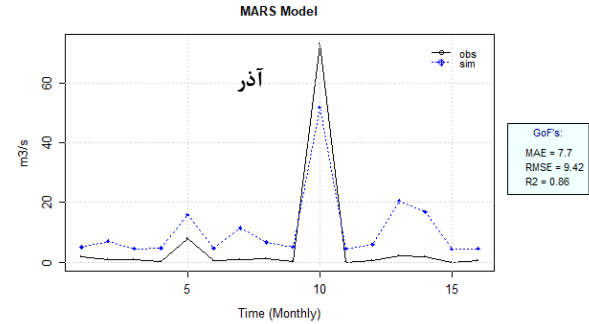
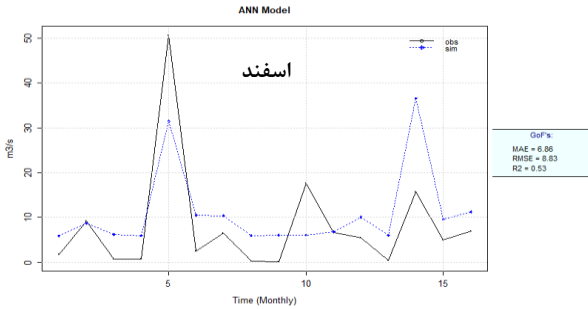
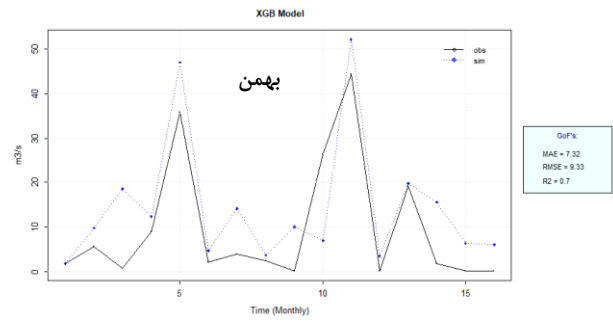
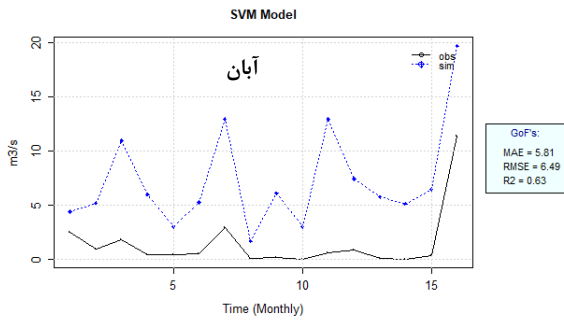
شکل ۵- نمودار پراکنش مقادیر مشاهده‌ای و شبیه‌سازی شده مدل‌های مورد استفاده در مرحله صحت‌سنجی

در مهر ماه است. اما در همین دوره زمانی معیارهای ارزیابی MAE و RMSE کمترین مقادیر خطا را داشته‌اند. در کل بر اساس معیارهای ارزیابی، بهترین عملکرد در بین مدل‌ها در دوره‌های زمانی مورد بررسی مربوط به مدل MARS با مقدار RMSE برابر با ۰/۶ مترمکعب در ثانیه، R^2 برابر با ۰/۸۸ و MAE برابر با ۵/۳۹ مترمکعب در ثانیه در بازه زمانی پاییز است. شکل (۶)، نمودار سری زمانی مقادیر مشاهده‌ای و شبیه‌سازی شده تعدادی از مدل‌های برتر در بازه‌های زمانی مرتبط با SST را نشان می‌دهد. همان‌طور که شکل (۶)، نشان می‌دهد نتایج مدل‌های شبیه‌سازی در این مطالعه داری بیش‌برآورد حدود ۵ مترمکعب در ثانیه است البته در دبی‌های اوج این بیش‌برآورد به شدت افزایش می‌یابد.

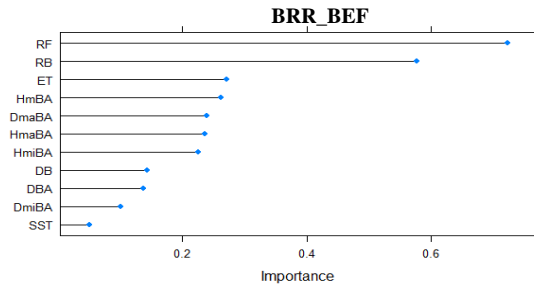
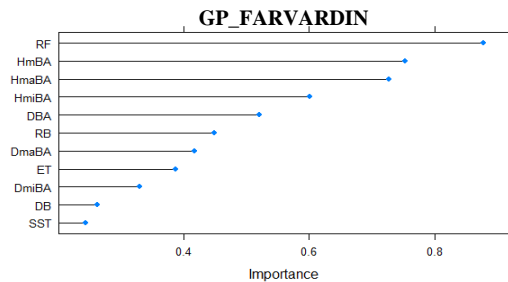
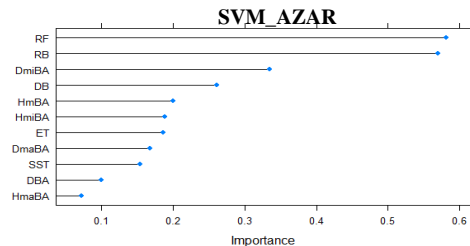
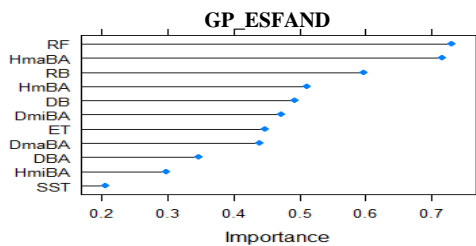
نتایج شبیه‌سازی دبی ماهانه در ارتباط با SST همان‌طور که اشاره گردید در ادامه با استفاده از نتایج مطالعه Mohammadi, (2020) به بررسی مؤثرترین سلول SST بر روی نتایج شبیه‌سازی جریان رواناب با استفاده از مدل‌های فوق‌الذکر پرداخته شد. جدول (۴)، نتایج مدل‌سازی و بهترین مدل را در هر بازه زمانی نشان می‌دهد. چنان‌که از جدول (۴) مشخص می‌شود در بازه‌های زمانی متفاوت، مدل‌های متفاوتی بهترین عملکرد را داشته‌اند و نمی‌توان یک مدل را برای تمامی بازه‌های زمانی معرفی نمود. بر اساس معیار ارزیابی R^2 بهترین عملکرد در بین مدل‌ها در بازه‌های زمانی متفاوت مربوط به مدل MARS با مقدار ۰/۸۸ در بازه زمانی مهر تا آذر است. ضعیف‌ترین عملکرد بر اساس این معیار ارزیابی، مربوط به مدل XGB با مقدار ۰/۱۴

جدول ۴- نتایج ارزیابی مدل‌ها و شناسایی مناسب‌ترین مدل در بازه‌های زمانی در ارتباط با SST

صحت‌سنجی			آموزش			روش	مخفف مدل	مناسب‌ترین مدل	دوره زمانی
MAE	RMSE	R^2	MAE	R^2	RMSE				
۳/۰۸	۰/۱۴	۳/۵۵	۲/۹۷	۰/۰۷	۴/۷۳	xgbLinear	XGB	الگوریتم‌های ارتقای شدید گرادیان	مهر
۵/۸۱	۰/۶۳	۶/۴۹	۴/۶۳	۰/۱۶	۷/۱۹	"svmLinear3	SVM	ماشین بردار پشتیبان	آبان
۷/۷	۰/۸۶	۹/۴۲	۶/۷۹	۰/۵۵	۱۰/۴۳	bagEarth	MARS	مدل اسپلاین چند متغیره رگرسیون انطباقی	آذر
۶/۲۶	۰/۵۷	۹/۸۳	۱۸/۲	۰/۵۸	۳۹/۰۴	"svmLinear3	SVM	ماشین بردار پشتیبان	دی
۷/۳۲	۰/۷	۹/۳۳	۱۹/۳۱	۰/۵۴	۳۳/۰۴	xgbLinear	XGB	الگوریتم‌های ارتقای شدید گرادیان	بهمن
۶/۸۶	۰/۵۳	۸/۸۳	۱۳/۹۴	۰/۵۴	۱۶/۳۶	brnn	ANN	شبکه عصبی مصنوعی	اسفند
۴/۹۷	۰/۳۱	۵/۶۹	۸/۹۴	۰/۵۹	۱۶/۶۳	gaussprRadial	GP	فرایند گوسی	فروردین
۵/۳۹	۰/۸۸	۶/۳	۴/۳۹	۰/۴۲	۸/۳۲	bagEarth	MARS	مدل اسپلاین چند متغیره رگرسیون انطباقی	مهر آبان آذر
۱۰/۸۵	۰/۵۵	۱۳/۸۳	۱۵/۹۲	۰/۵۵	۳۰/۵۷	'cforest'	RF	الگوریتم جنگل تصادفی	دی بهمن اسفند
۷/۸۲	۰/۵۹	۱۰/۳۸	۱۲/۴۳	۰/۵۵	۲۴/۲۷	xgbLinear	XGB	الگوریتم‌های ارتقای شدید گرادیان	بهمن اسفند فروردین



شکل ۶- نمودار سری زمانی مقادیر مشاهده‌ای و شبیه‌سازی شده تعدادی از مدل‌های برتر در بازه‌های زمانی مختلف



شکل ۷- نمودار اهمیت نسبی پارامترهای ورودی تعدادی از مدل‌ها

RF= باران فاریاب، RB= باران برطین، ET= تبخیر و تعرق برطین، HmiBA= رطوبت متوسط بندرعباس، HmaBA= رطوبت حداکثر بندرعباس، HmiBA= رطوبت حداقل بندرعباس، DBA= دمای متوسط بندرعباس، DmiBA= دمای حداقل بندرعباس، DmaBA= دمای حداکثر بندرعباس، SST= دمای سطح دریا

نتایج شبیه‌سازی مدل‌ها در بازه‌های زمانی مورد اشاره، نتایج نشان داد که SST نه تنها موجب بهبود نتایج شبیه‌سازی رواناب در

با مقایسه جدول‌های (۳ و ۴) مشخص می‌گردد علی‌رغم این‌که انتظار می‌رفت، دخیل نمودن مقادیر SST موجب بهبود

۷/۷ مترمکعب در ثانیه در دوره صحت‌سنجی کاهش داشته است. این نتایج با نتایج Samadi et al (2019) و Mirshekari et al (2019) که در مطالعات آن‌ها نیز عملکرد مدل در مرحله صحت‌سنجی بهتر از مرحله آموزش بوده است همخوانی دارد.

چنان‌که اشاره شد در این مطالعه، مقادیر SST مؤثر که توسط مطالعه Mohammadi, (2020) تعیین گردیده بود نیز به‌منظور پاسخ به این سوال که آیا وارد نمودن مقادیر SST به‌عنوان یک ورودی در مدل‌سازی بارش- رواناب آبخیز میناب موجب بهبود نتایج شبیه‌سازی می‌گردد یا خیر؟ مورد ارزیابی و بررسی قرار گرفت. نتایج نشان داد که پاسخ این سؤال منفی است، به‌طوری‌که مقادیر SST موجب بهبود نتایج شبیه‌سازی رواناب در منطقه مورد مطالعه نگردید. در بعضی ماه‌ها مانند مهر و فروردین موجب کاهش چشم‌گیر مقادیر R^2 گردید. نتایج بررسی اهمیت نسبی پارامترهای ورودی مدل‌ها در این مرحله نیز نشان داد که مقادیر SST کمترین اهمیت نسبی را در بین ورودی‌های مدل‌ها دارا است. در مجموع می‌توان بیان داشت که مقادیر SST موجب بهبود نتایج شبیه‌سازی رواناب در آبخیز میناب نمی‌گردد. با توجه به آنکه Nazemosadat et al (1995) و Nazemosadat (1998) در مطالعات خود روشن ساختند که نوسان‌های دمای سطح آب خلیج فارس نشان معنی‌داری بر نوسان‌های بارش در بسیاری از گستره‌های جنوب غربی و جنوبی ایران دارد، اما این مطالعه نشان داد که در مقیاس کوچک و در سطح آبخیز تاثیر معنی‌داری بر شبیه‌سازی بارش رواناب ندارد از طرف دیگر این موضوع نشان‌دهنده ماهیت پیچیده فرآیندهای بارش رواناب است که در تبدیل بارش به رواناب پارامترهای فیزیکی بسیار زیادی دخیل می‌باشند.

"هیچ‌گونه تعارض منافع بین نویسندگان وجود ندارد"

منطقه مورد مطالعه نگردید؛ بلکه در بعضی مواقع همچون مهرماه و فروردین‌ماه موجب کاهش چشم‌گیر مقادیر R^2 گردیده است. لذا، در ادامه به‌منظور تعیین جایگاه و اهمیت نسبی مقادیر SST به‌عنوان یک ورودی در بین ورودی‌های مدل با استفاده از تابع varImp بسته (Kuhn, 2012 R caret) پرداخته شد. شکل (۷)، نتایج اهمیت نسبی پارامترهای ورودی تعدادی از مدل‌ها را نشان می‌دهد. چنان‌که شکل (۷) نشان می‌دهد در بین ورودی‌های مدل‌های شبیه‌سازی جریان در منطقه مورد مطالعه، مقادیر SST کمترین اهمیت نسبی را دارد، که مؤید این موضوع است که مقادیر SST موجب بهبود نتایج شبیه‌سازی رواناب در منطقه مورد مطالعه نمی‌شوند.

نتیجه‌گیری

در این مطالعه به‌منظور شبیه‌سازی بارش- رواناب آبخیز میناب از هشت مدل شبیه‌سازی بر پایه یادگیری ماشین استفاده گردید و نتایج با استفاده از روش‌های ارزیابی و همچنین نمودار تیلور مورد ارزیابی قرار گرفت و نتایج نشان داد که مدل MARS بهترین عملکرد را در بین مدل‌ها برای شبیه‌سازی دبی ماهانه آبخیز مورد مطالعه داشته است. مدل SVM نیز با مقدار RSME برابر ۷/۷۳ مترمکعب در ثانیه عملکرد مناسبی داشته است. بقیه مدل‌ها نیز عملکرد نسبتاً نزدیک به هم داشته‌اند، به‌طوری‌که مدل XGB با مقدار ۹/۹۸ مترمکعب در ثانیه بالاترین و مدل MARS با مقدار ۷/۷ مترمکعب در ثانیه، کمترین مقدار RMSE را داشته‌اند. همچنین، عملکرد مدل‌ها در دوره صحت‌سنجی نسبت به دوره آموزش بهبود بسیاری خوبی داشته‌اند، به‌طوری‌که به‌طور نمونه مقادیر RMSE مدل MARS از مقدار ۲۰/۴۱ مترمکعب در ثانیه برای آموزش به مقدار

REFERENCES

- Ahmadi, F. (2019). Evaluation of Support Vector Machine and Adaptive Neuro-Fuzzy Inference System Performance in Prediction of Monthly River Flow (Case Study: Nazlu chai and Sezar Rivers). *Iranian Journal of Soil and Water Research*, 51(3), 686-673.
- Aleotti, P., and Chowdhury, R. (1999). Landslide hazard assessment: summary review and new perspectives. *Bulletin of Engineering Geology and the environment*, 58(1), 21-44.
- Ångström, A., (1935). Teleconnections of Climatic Changes in Present Time. *Geografiska Annaler*, 17: 242-258.
- Bayat varkashi, M. and Gheysari, P. (2018). The Effect of Enso Phenomenon on Groudwater Table (Case Study: Esfahan and Hormozgan). *Iran-Water Resources Research*, 14 (2), 1-15. (In Farsi)
- Bergström, S., Carlsson, B., Gardelin, M., Lindström, G., Pettersson, A. and Rummukainen, M. (2001). Climate change impacts on runoff in Sweden - assessments by global climate models, dynamical downscaling and hydrological modelling. *Climate Research*, 16(2), 101-112.
- Bhattacharya, B., and Solomatine, D. P. (2005). Neural networks and M5 model trees in modelling water level-discharge relationship. *Neurocomputing*, 63, 381-396.
- Breiman, L. (2001). Random forests. *Machine learning*, 45(1), 5-32.
- Bui, D. T., Pradhan, B., Lofman, O., Revhaug, I., and Dick, O. B. (2012). Landslide susceptibility mapping at Hoa Binh province (Vietnam) using an adaptive neuro-fuzzy inference system and GIS. *Computers and Geosciences*, 45, 199-211.
- Chemura, A., Rwasoka, D., Mutanga, O., Dube, T., and Mushore, T. (2020). The impact of land-use/land cover changes on water balance of the heterogeneous Buzi sub-catchment, Zimbabwe. www.sid.ir

Remote Sensing Applications: Society and Environment, 18, 100292.

- Chen, T., and Guestrin, C. (2016, August). Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining* (pp. 785-794).
- Chen, T., Wang, X., Chu, Y., Wei, D.-Q., and Xiong, Y. (2020,a). T4SEXGB: interpretable sequence-based prediction of type IV secreted effectors using extreme gradient boosting algorithm. *bioRxiv*. doi: 10.1101/2020.06.18.158253
- Chen, W., Li, Y., Xue, W., Shahabi, H., Li, S., Hong, H. and Ahmad, B. B. (2020,b). Modeling flood susceptibility using data-driven approaches of naïve bayes tree, alternating decision tree, and random forest methods. *Science of the Total Environment*, 701, 134979.
- Chiew, F.H.S., Piechota, T.C., Dracup, J.A. and McMahon, T.A., (1998). El Nino/southern oscillation and Australian rainfall, stream flow and drought: links and potential for forecasting. *Journal of Hydrology*, 204(1-4): 138-149.
- Dastorani, M. T., Mahjoobi, J., Talebi, A., and Fakhar, F. (2018). Application of Machine Learning Approaches in Rainfall-Runoff Modeling (Case Study: Zayandeh Rood Basin in Iran). *Civil Engineering Infrastructures Journal*, 51(2), 293-310.
- Drobinski, P., Silva, N.D., Panthou, G., Bastin, S., Muller, C., Ahrens, B., Borga, B., Conte, D., Fosser, G., Giorgi, F., Güttler, I., Kotroni, V., Li, L., Morin, E., Öno, B., Quintana-Segui, P., Romera, R., and Torma, C. S. (2018). Scaling precipitation extremes with temperature in the Mediterranean: past climate assessment and projection in anthropogenic scenarios, *Clim. Dynam.*, 51, 1237-1257.
- Eini, M., Kaboli, H. S., Rashidian, M., and Hedayat, H. (2020). Hazard and vulnerability in urban flood risk mapping: Machine learning techniques and considering the role of urban districts. *International Journal of Disaster Risk Reduction*, 101687.
- Etemad-Shahidi, A., and Bonakdar, L. (2009). Design of rubble-mound breakwaters using M5' machine learning method. *Applied Ocean Research*, 31(3), 197-201.
- Etemad-Shahidi, A., and Taghipour, M. (2012). Predicting longitudinal dispersion coefficient in natural streams using M5' model tree. *Journal of hydraulic engineering*, 138(6), 542-554.
- Evans, M.N., Fairbanks, R.G. and Rubenstone, J.L. (1998). A proxy index of ENSO Teleconnections. *Nature*, 394(6695): 732-733.
- Farmer, W. H., and Vogel, R. M. (2016). On the deterministic and stochastic use of hydrologic models. *Water Resources Research*, 52(7), 5619-5633.
- Fausett, L. (1994). Fundamentals of neural networks: architectures, algorithms, and applications. Prentice-Hall, Inc.
- Friedman, J. H. (1991). Multivariate adaptive regression splines. *The annals of statistics*, 1-67.
- Friedman, J. H. (2001). Greedy function approximation: a gradient boosting machine. *Annals of statistics*, 1189-1232.
- Gholami, H., Mohamadifar, A., and Collins, A. L. (2020). Spatial mapping of the provenance of storm dust: Application of data mining and ensemble modelling. *Atmospheric Research*, 233, 104716.
- Ghorbani, K., Sohrabian, E. and Salarijazi, M. (2016). Evaluation of Hydrological and Data Mining Models in Monthly River Discharge Simulation and Prediction (Case Study: Araz-Kouseh Watershed). *Journal of Water and Soil Conservation*, 23(1), 203-217. (In Farsi)
- Gramacy, R. B., (2019). Package 'monomvn'. Package 'monomvm' version 1.9-13. Retrieved from <https://cran.r-project.org/web/packages/momomvm/index.html>
- Granata, F. (2019). Evapotranspiration evaluation models based on machine learning algorithms—A comparative study. *Agricultural Water Management*, 217, 303-315.
- Guan, H., He, X. and Zhang, X. (2015). A comprehensive examination of global atmospheric CO2 teleconnections using wavelet-based multi-resolution analysis. *Environmental Earth Sciences*, 74(10): 7239-7253.
- He, X. and Guan, H. (2013). Multiresolution analysis of precipitation teleconnections with large-scale climate signals: A case study in South Australia. *Water Resources Research*, 49(10), 6995-7008.
- Helleputte, T., and Gramme, P. (2017). LiblineaR: Linear predictive models based on the LIBLINEAR C/C++ Library. *R package version*, 2-10.
- Hornik, K., Buchta, C., Hothorn, T., Karatzoglou, A., Meyer, D., Zeileis, A., & Hornik, M. K. (2020). Package 'RWeka' version 0.4-43. Retrieved from <https://cran.r-project.org/web/packages/RWeka/index.html>.
- Hosseini, F., Karimi, O. and Hamedi, F. (2019). Survival Analysis using Bayesian Additive Regression Trees. *Andishe*, 24(1), 33-42. (In Farsi)
- Hrachowitz, M., Savenije, H. H. G., Blöschl, G., McDonnell, J. J., Sivapalan, M., Pomeroy, J. W., ... and Fenicia, F. (2013). A decade of Predictions in Ungauged Basins (PUB)—a review. *Hydrological sciences journal*, 58(6), 1198-1255.
- Iorgulescu, I., and Beven, K. J. (2004). Nonparametric direct mapping of rainfall-runoff relationships: An alternative approach to data analysis and modeling?. *Water Resources Research*, 40(8).
- Jansa, A., Alpert, P., Arbogast, P., Buzzi, A., Ivancan-Picek, B., Kotroni, V., Llasat, M. C., Ramis, C., Richard, E., Romero, R., and Speranza. (2014). A.: MEDEX: a general overview, *Nat. Hazards*

- Earth Syst. Sci.*, 14, 1965–1984.
- Kapelner, A., and Bleich, J. (2013). bartMachine: Machine learning with Bayesian additive regression trees. *arXiv preprint arXiv:1312.2171*.
- Karatzoglou, A., Smola, A., Hornik, K., and Karatzoglou, M. A. (2019). Package ‘kernlab’. CRAN R Project.
- Khosravi, K., Cooper, J. R., Daggupati, P., Pham, B. T., and Bui, D. T. (2020). Bedload transport rate prediction: Application of novel hybrid data mining techniques. *Journal of Hydrology*, 124774.
- Kirono, D.G.C., Chiew, F.H.S. and Kent, D.M. (2010). Identification of best predictors for forecasting seasonal rainfall and runoff in Australia. *Hydrological Processes*, 24(10): 1237–1247.
- Kumar, A., Kumar, P., and Singh, V. K. (2019). Evaluating Different Machine Learning Models for Runoff and Suspended Sediment Simulation. *Water resources management*, 33(3), 1217-1231.
- Liaw, A., and Wiener, M. (2002). Classification and regression by randomForest. *R news*, 2(3), 18-22.
- Luce, C. (2014). Runoff Prediction in Ungauged Basins: Synthesis Across Processes, Places and Scales: Edited by Günter Blöschl, Murugesu Sivapalan, Thorsten Wagener, Alberto Viglione, and Hubert Savenije Cambridge University Press, 2013, 465 pp., ISBN: 978-1107028180, 140(hardback), 112 (eBook). *Eos, Transactions American Geophysical Union*, 95(2), 22-22.
- Marjanović, M., Kovačević, M., Bajat, B., and Voženilek, V. (2011). Landslide susceptibility assessment using SVM machine learning algorithm. *Engineering Geology*, 123(3), 225-234.
- Mekanik, F., Imteaz, M.A. and Talei, A., (2015). Seasonal rainfall forecasting by adaptive network-based fuzzy inference system (ANFIS) using large scale climate signals. *Climate Dynamics*, 46(9-10): 3097-3111
- Milborrow, S. (2020). Package ‘earth version 5.3. Retrieved from rproject.org/web/packages/earth/index.html
- Mohammadi, M. (2020). Development of Conjunctive Use Model of Surface and Groundwater Based on Teleconnection Patterns Forecasts to Study Ground Water Balancing Scenarios. (Case study: downstream of Minab Esteghlal Dam. Ph. D. dissertation, University of Hormozgan, BandarAbbas, Iran. (In Farsi)
- Nazemosadat, M. J., (1998). The Persian Gulf sea surface temperature as a drought diagnostic for southern parts of Iran, Drouth News Network, 10, 12-14.
- Nazemosadat, M. J., Cordery, I., and Eslamian, S., (1995). The impact of the Persian Gulf Sea surface temperature on Iranian rainfall, Proceedings of the Iranian Water Resource Management Conference, Esfahan, Iran, 809-819.
- Nazemosadat, M. J., Ghaedamini, H. and Tavakoli, M. (2014). Investigating the climate change diagnostics over the north western parts of the Indian Ocean: The SST analysis for the period 1950-2009. *Iranian Journal of Geophysics*, 8(2), 26-40.
- Niu, W. J., Feng, Z. K., Zeng, M., Feng, B. F., Min, Y. W., Cheng, C. T., and Zhou, J. Z. (2019). Forecasting reservoir monthly runoff via ensemble empirical mode decomposition and extreme learning machine optimized by an improved gravitational search algorithm. *Applied Soft Computing*, 82, 105589.
- Nobre, J., and Neves, R. F. (2019). Combining principal component analysis, discrete wavelet transform and XGBoost to trade in the financial markets. *Expert Systems with Applications*, 125, 181-194.
- Nohegar, A., Ghashghaeizadeh, N., Heydarzadeh, M., Eydoon, M. and Pannahi, M. (2016). Assessment of drought and its impact on surface and groundwater resources (Case study: River basin Minab), *Journal of Earth Science Researches*, 7(27), 28-43. (In Farsi)
- Nohegar, A., Torabi, B., Holisaz, A. and Biniyaz, M. (2013). Soil Erosion Model Implication (Case Study: Minab Basin). *E.E.R*, 3(2), 53-64. (In Farsi)
- Nourani, V., Davanlou Tajbakhsh, A., Molajou, A., and Gokcekus, H. (2019). Hybrid Wavelet-M5 Model tree for rainfall-runoff modeling. *Journal of Hydrologic Engineering*, 24(5), 04019012.
- O'Hagan, A. (1978). Curve fitting and optimal design for prediction. *Journal of the Royal Statistical Society: Series B (Methodological)*, 40(1), 1-24.
- Panahi, M., Gayen, A., Pourghasemi, H. R., Rezaie, F., and Lee, S. (2020). Spatial prediction of landslide susceptibility using hybrid support vector regression (SVR) and the adaptive Neuro-fuzzy inference system (ANFIS) with various metaheuristic algorithms. *Science of the Total Environment*, 139937.
- Peddle, D. R., Foody, G. M., Zhang, A., Franklin, S. E., and LeDrew, E. F. (1994). Multi-source image classification II: An empirical comparison of evidential reasoning and neural network approaches. *Canadian Journal of Remote Sensing*, 20(4), 396-407.
- Peters, D.L., Atkinson, D., Monk, W.A., Tenenbaum, D.E. and Baird, D.J., (2013). A multi-scale hydroclimatic analysis of runoff generation in the Athabasca River, western Canada. *Hydrological Processes*, 27(13), 1915-1934.
- Pourghasemi, H. R., and Rahmati, O. (2018). Prediction of the landslide susceptibility: Which algorithm, which precision?. *Catena*, 162, 177-192.
- Pourghasemi, H.R., Yousefi, S., Kornejady, A., Cerda, A. (2017). Performance assessment of individual and ensemble data-mining techniques for gully erosion modeling. *Sci. Total Environ.* 609, 764–775.
- Quinlan, J. R. (1992, November). Learning with continuous classes. In *5th Australian joint conference on artificial intelligence* (Vol. 92, pp. 343-348).
- Rasmussen, C. E., and Williams, C. K. (2006). *Gaussian Processes for Machine Learning* the MIT Press. Cambridge, MA.
- Razavi, T., and Coulibaly, P. (2013) www.sid.ir

- prediction in ungauged basins: review of regionalization methods. *Journal of hydrologic engineering*, 18(8), 958-975.
- Riad, S., Mania, J., Bouchaou, L., Najjar, Y., (2004). Rainfall-runoff model using an artificial neural network approach. *Math. Comput. Modell.* 40 (7–8), 839–846.
- Rodriguez-Galiano, V. F., Ghimire, B., Rogan, J., Chica-Olmo, M., and Rigol-Sanchez, J. P. (2012). An assessment of the effectiveness of a random forest classifier for land-cover classification. *ISPRS Journal of Photogrammetry and Remote Sensing*, 67, 93-104.
- Rodriguez-Galiano, V., Sanchez-Castillo, M., Chica-Olmo, M., and Chica-Rivas, M. J. O. G. R. (2015). Machine learning predictive models for mineral prospectivity: An evaluation of neural networks, random forest, regression trees and support vector machines. *Ore Geology Reviews*, 71, 804-818.
- Rodriguez-Galiano, V., and Gianola, D. (2020). Package 'brnn' version 0.8. Retrieved from <https://cran.r-project.org/web/packages/brnn/index.html>
- Samadi, M., Bahremand, A., and Fathabadi, A. (2019). The Boustan Dam monthly inflow forecasting using data-driven and ensemble models in the Golestan Province. *Journal of Watershed Engineering and Management*, 11(4), 1044-1058. (In Farsi)
- Samadi, M., Jabbari, E., Azamathulla, H. M., and Mojallal, M. (2015). Estimation of scour depth below free overfall spillways using multivariate adaptive regression splines and artificial neural networks. *Engineering Applications of Computational Fluid Mechanics*, 9(1), 291-300.
- Sattari, M., Pourazad, M. and Najafabadi, R. (2016). Technical Note: Hourly River flow forecast of Aharchay River using machine learning methods. *Journal of Watershed Engineering and Management*, 8(1), 115-127. (In Farsi)
- Seeger, M. (2004). Gaussian processes for machine learning. *International journal of neural systems*, 14(02), 69-106.
- Sezen, C., Bezak, N., Bai, Y., and Šraj, M. (2019). Hydrological modelling of karst catchment using lumped conceptual and data mining models. *Journal of Hydrology*, 576, 98-110.
- Shawe-Taylor, J., and Cristianini, N. (2000). An introduction to support vector machines.
- Shortridge, J. E., Guikema, S. D., and Zaitchik, B. F. (2016). Machine learning methods for empirical streamflow simulation: a comparison of model accuracy, interpretability, and uncertainty in seasonal watersheds. *Hydrology and Earth System Sciences*, 20(7).
- Sobolowski, S. and Frei, A. (2007). Lagged relationships between North American snow mass and atmospheric teleconnection indices. *International Journal of Climatology*, 27(2): 221-231.
- Taylor, K. E. (2001). Summarizing multiple aspects of model performance in a single diagram. *Journal of Geophysical Research: Atmospheres*, 106(D7), 7183-7192.
- Vapnik, V. (1998). *Statistical Learning Theory* John Wiley. *New York*.
- Vapnik, V. N. (2000). Direct methods in statistical learning theory. In *The nature of statistical learning theory* (pp. 225-265). Springer, New York, NY.
- Wang, J., Wang, X., Lei, X-h., Wang, H., Zhang, X-h., Tan, Q-f., Liu, X-l. (2019). Teleconnection analysis of monthly streamflow using ensemble empirical mode decomposition, *Journal of Hydrology*, (19)31146
- Wang, Y., and Witten, I. H. (1996). Induction of model trees for predicting continuous classes.
- Worland, S. C., Farmer, W. H., and Kiang, J. E. (2018). Improving predictions of hydrological low-flow indices in ungauged basins using machine learning. *Environmental modelling and software*, 101, 169-182.
- Wu, X., Ren, F., and Niu, R. (2014). Landslide susceptibility assessment using object mapping units, decision tree, and support vector machine models in the Three Gorges of China. *Environmental earth sciences*, 71(11), 4725-4738.
- Yang, C. C., Prasher, S. O., Lacroix, R., and Kim, S. H. (2003). A multivariate adaptive regression splines model for simulation of pesticide transport in soils. *Biosystems Engineering*, 86(1), 9-15.
- Yaseen, Z. M., El-Shafie, A., Jaafar, O., Afan, H. A., and Sayl, K. N. (2015). Artificial intelligence based models for stream-flow forecasting: 2000–2015. *Journal of Hydrology*, 530, 829-844.
- Yaseen, Z. M., Kisi, O., and Demir, V. (2016). Enhancing long-term streamflow forecasting and predicting using periodicity data component: application of artificial intelligence. *Water resources management*, 30(12), 4125-4151.
- Zahiri, J., and Nezaratian, H. (2020). Estimation of transverse mixing coefficient in streams using M5, MARS, GA, and PSO approaches. *Environmental Science and Pollution Research*, 1-14.
- Zema, D. A., Lucas-Borja, M. E., Fotia, L., Rosaci, D., Sarnè, G. M., and Zimbone, S. M. (2020). Predicting the hydrological response of a forest after wildfire and soil treatments using an Artificial Neural Network. *Computers and Electronics in Agriculture*, 170, 105280.