Gorgan University of Agricultural Sciences and Natural Resources

Journal of Water and Soil Conservation

Print ISSN: 2322-2069
Online ISSN: 2322-2794

Monthly and seasonal runoff estimation using time series, decision tree, and multivariable linear regression

Hedieh Khodakhah[1] | Khalil Ghorbani[*2] | Meysam Salarijazi[3] | Mohamamd Abdolhosseini[4]

1. M.Sc. Student, Dept. of Water Resources Engineering, Gorgan University of Agricultural Sciences and Natural Resources, Gorgan, Iran. E-mail: hedieh.khodakhah_s98@gau.ac.ir
2. Corresponding Author, Associate Prof., Dept. of Water Engineering, Gorgan University of Agricultural Sciences and Natural Resources, Gorgan, Iran. E-mail: ghorbani.khalil@gmail.com
3. Associate Prof., Dept. of Water Engineering, Gorgan University of Agricultural Sciences and Natural Resources, Gorgan, Iran. E-mail: meysam.salarijazi@gmail.com
4. Assistant Prof., Dept. of Water Engineering, Gorgan University of Agricultural Sciences and Natural Resources, Gorgan, Iran. E-mail: abd.phd@gmail.com

Article Info

Article type:
Research Full Paper

Article history:
Received: 01.21.2022
Revised: 03.11.2022
Accepted: 03.16.2022

Keywords:
Decision tree M5,
Multivariable linear regression,
Navrood watershed,
Runoff,
Time series

ABSTRACT

Background and Objectives: One of the essential factors in the programming and management of water resources is predicting the amount of runoff. Increasing the accuracy in predicting runoff will increase the efficiency of programming and management; therefore, improving the modeling of discharge prediction is a requisite issue. The first aim of this study is to evaluate the efficiency of the multivariable linear regression, M5 decision tree, and time series in predicting the river runoff. The second aim is to analyze the modeling time step (monthly or seasonal) and the effects of model inputs (one delay steps variable against several delay steps variable) on the accuracy of the studied models.

Material and Methods: Navrood watershed located in the west part of Guilan province is chosen for the study area in this research. Required data is collected from Kharjgil (1368-1398) and Kholian (1375-1397), including monthly river flow, rainfall, and temperature from Guilan regional water company. The amount of runoff is predicted in two approaches by the received data in monthly and seasonal time steps sing three models of multivariable linear regression, time series, and M5 decision tree. In the first approach, input variables to the model were river flow, rainfall, and temperature with three steps delay. In the second approach, the only variable was river flow with three steps delay. The model evaluation criteria in this research are the mean bias error (MBE), Nash-Sutcliffe efficiency (NSE), and coefficient of determination (R^2).

Results: In the first approach and in monthly timestep, M5 decision tree is selected model with MBE-NSE equal to -0.04,0.80 (train) and 0.01,0.72 (test) in Kharjgil station, and -0.01,0.79 (train) and 0.00,0.86 (test) in Kholian station. In the seasonal time step, the criteria for the M5 decision tree in Kholian station are equal to 0.02,0.78 (train), -0.02,0.86 (test), and in Kholian station are -0.01,0.79 (train), 0.00,0.86 (test). This model was the best in this study for the first approach in the seasonal time step. The second approach has led to different findings considering both monthly and seasonal time steps. In the second approach, the criteria in monthly time step for time series model during train and test in Kharjgil station are

respectively -0.05,0.47 and 0.10,0.52 and in Kholian are -0.02,0.63 and 0.2,0.49. The selected model criteria for seasonal time step considering train and test are -0.42,0.58 and 0.06,0.83 in Kharjgil station, and 0.09,0.40 and -0.10,0.62 in Kholian station. The time series model is selected in the second approach in the seasonal time step.

Conclusion: The findings of this research have shown that in both stations and time steps, the M5 decision tree model has shown a higher accuracy in prediction than the two other models in the first approach. Meanwhile, the decision tree model does not show accurate results in the second approach. Alternatively, compared to two other models in both stations and both time steps, the time series model had a higher accuracy. Findings of this research have emphatically shown that specific approaches in choosing the model's inputs can effectively influence the selected model and the accuracy of modeling.

Cite this article: Khodakhah, Hedieh, Ghorbani, Khalil, Salarijazi, Meysam, Abdolhosseini, Mohamamd. 2022. Monthly and seasonal runoff estimation using time series, decision tree, and multivariable linear regression. *Journal of Water and Soil Conservation*, 28 (4), 27-52.

© The Author(s).
DOI: 10.22069/jwsc.2022.19921.3533
Publisher: Gorgan University of Agricultural Sciences and Natural Resources

نشریه پژوهش‌های حفاظت آب و خاک

شاپا چاپی: ۲۰۶۹-۲۳۲۲
شاپا آنلاین: ۲۷۹٤-۲۳۲۲

برآورد رواناب ماهانه و فصلی با مدل‌های سری زمانی، درخت تصمیم و رگرسیون خطی چندمتغیره

هدیه خداخواه[۱] | خلیل قربانی*[۲] | میثم سالاری‌جزی[۳] | محمد عبدالحسینی[٤]

۱. دانشجوی کارشناسی‌ارشد گروه مهندسی منابع آب، دانشگاه علوم کشاورزی و منابع طبیعی گرگان. رایانامه: hedieh.khodakhah_s98@gau.ac.ir
۲. نویسنده مسئول، دانشیار گروه مهندسی آب، دانشگاه علوم کشاورزی و منابع طبیعی گرگان. رایانامه: ghorbani.khalil@gmail.com
۳. دانشیار گروه مهندسی آب، دانشگاه علوم کشاورزی و منابع طبیعی گرگان. رایانامه: meysam.salarijazi@gmail.com
٤. استادیار گروه مهندسی آب، دانشگاه علوم کشاورزی و منابع طبیعی گرگان. رایانامه: abd.phd@gmail.com

اطلاعات مقاله

نوع مقاله:
مقاله کامل علمی- پژوهشی

تاریخ دریافت: ۱٤۰۰/۱۱/۰۱
تاریخ ویرایش: ۱٤۰۰/۱۲/۲۰
تاریخ پذیرش: ۱٤۰۰/۱۲/۲۵

واژه‌های کلیدی:
پیش‌بینی،
حوزه آبخیز ناورود،
درخت تصمیم‌گیری M5،
رگرسیون چندمتغیره خطی،
رواناب،
سری زمانی

چکیده

سابقه و هدف: ازجمله عوامل دارای اهمیت در مدیریت و برنامه‌ریزی منابع آب پیش‌بینی مقدار رواناب است. با افزایش دقت در پیش‌بینی رواناب رودخانه مدیریت و برنامه‌ریزی کارآمدتری صورت می‌گیرد بنابراین بهبود مدل‌سازی پیش‌بینی رواناب امری ضروری است. اولین هدف از این مطالعه ارزیابی کارایی مدل‌های رگرسیون چندمتغیره خطی، درخت تصمیم M5 و سری زمانی در پیش‌بینی رواناب رودخانه است. هدف دوم بررسی مقیاس زمانی مدل‌سازی (ماهانه و فصلی) و نیز تأثیر ورودی‌های مدل (یک متغیر با گام‌های تأخیر و چند متغیر با گام‌های تأخیر) بر دقت مدل‌های موردمطالعه است.

مواد و روش‌ها: در این پژوهش حوزه آبخیزناورود واقع در غرب استان گیلان جهت منطقه مطالعاتی انتخاب‌شده است. داده‌های موردنیاز دو ایستگاه خرجگیل در سال‌های ۱۳۹۸-۱۳٦۸ و خلیان در سال‌های ۱۳۹۷-۱۳۷۵ شامل دبی، بارش و دما در مقیاس زمانی ماهانه از آب منطقه‌ای استان گیلان جمع‌آوری‌شده است. مقدار رواناب توسط داده‌های دریافت شده در بازه زمانی ماهانه و فصلی با استفاده سه مدل رگرسیون چندمتغیره خطی، سری زمانی و درخت تصمیم M5 در دو رویکرد متفاوت پیش‌بینی‌شده است. رویکرد اول متغیرهای ورودی به مدل شامل دبی، بارش و دما با ۳ گام تأخیر زمانی و در رویکرد دوم تنها متغیر دبی با ۳ گام تأخیر زمانی بوده است. شاخص‌های ارزیابی در این پژوهش شامل میانگین انحراف خطا (MBE)، ضریب کارایی مش (NSE) و ضریب تعیین (R^2) است.

یافته‌ها: در رویکرد اول و در پنجره زمانی ماهانه مدل درخت تصمیم M5 با شاخص MBE، NSE ۰/۰٤-، ۰/۸۰ (آموزش) و ۰/۰۱، ۰/۷۲ (آزمون) در ایستگاه خرجگیل و ۰/۰۱-، ۰/۷۹

(آموزش) و ۰/۰۰، ۰/۸۲ (آزمون) در ایستگاه خلیان به‌عنوان مدل منتخب انتخاب می‌گردد. در گام زمانی فصلی نیز مقادیر شاخص‌ها برای مدل درخت تصمیم M5 در ایستگاه خرجگیل برابر ۰/۰۲، ۰/۷۸ (آموزش) ۰/۰۲-، ۰/۸٦ (آزمون) و در ایستگاه خلیان نیز ۰/۰۱-، ۰/۷۹ (آموزش) و ۰/۰۰، ۰/۸٦ (آزمون) است و این مدل در گام زمانی فصلی در رویکرد اول نیز بهترین مدل موردمطالعه بوده است. رویکرد دوم در هر دو گام زمانی ماهانه و فصلی منجر به یافته‌های متفاوتی شده است. در رویکرد دوم در گام زمانی ماهانه مقادیر شاخص‌ها برای مدل سری زمانی در دو مرحله آموزش و آزمون در ایستگاه خرجگیل به ترتیب برابر ۰/۰۵-، ۰/٤۷ و ۰/۱۰، ۰/۵۲ و در ایستگاه خلیان برابر با ۰/۰۲-، ۰/٦۳ و ۰/۰۲، ۰/٤۹ بوده است. در گام زمانی فصلی نیز مقادیر شاخص‌های مدل منتخب در دو مرحله آموزش و آزمون در ایستگاه خرجگیل ۰/٤۲-، ۰/۵۸ و ۰/۰٦، ۰/۸۳ و خلیان ۰/۰۹، ۰/٤۰ و ۰/۱۰-، ۰/٦۲ است. در گام زمانی فصلی نیز مدل سری زمانی مدل منتخب در رویکرد دوم است.

نتیجه‌گیری: نتایج حاصل از این پژوهش بیانگر آن است که در رویکرد اول در هر دو ایستگاه و در هر دو گام زمانی مدل درخت تصمیم M5 دقت بالاتری در پیش‌بینی نسبت به دو مدل دیگر از خود نشان داده است درحالی‌که در رویکرد دوم مدل درخت تصمیم نتایج با دقت بالا از خود نشان نمی‌دهد و در مقابل مدل سری زمانی دقت بالاتری نسبت به دو مدل دیگر در هر دو ایستگاه و هر دو گام زمانی داشته است. یافته‌های این مطالعه بر این موضوع تأکید دارد که رویکرد مورداستفاده در انتخاب ورودی‌های مدل می‌تواند به شکل کامل مؤثری دقت مدل‌سازی و مدل منتخب را تحت‌تأثیر قرار دهد.

استناد: خداخواه، هدیه، قربانی، خلیل، سالاری‌جزی، میثم، عبدالحسینی، محمد (۱۴۰۰). برآورد رواناب ماهانه و فصلی با مدل‌های سری زمانی، درخت تصمیم و رگرسیون خطی چندمتغیره. *پژوهش‌های حفاظت آب و خاک*، ۲۸ (۴)، ۵۲-۲۷.

DOI: 10.22069/jwsc.2022.19921.3533

ناشر: دانشگاه علوم کشاورزی و منابع طبیعی گرگان

© نویسندگان.

مقدمه

عدم کنترل و پیش‌بینی مناسب رواناب در سال‌های اخیر یکی از عواملی بوده است که در هنگام وقوع سیلاب خسارات و تهدیداتی را با خود به همراه داشته است. به همین دلیل جهت پیشگیری از این خسارات احتمالی ناشی از سیلاب‌های ناگهانی و کنترل و مهار آن‌ها، پیش‌بینی رواناب امری ضروری و لازم‌الاجرا است چراکه بااطلاع از آن در هنگام وقوع سیل می‌توان اقدامات لازم جهت جلوگیری از خسارات ناشی از آن را به عمل آورد (۳۱). با توجه به اهمیت رواناب در مسائل مهندسی و مدیریت منابع سطحی، دستگاه‌های هشداردهنده سیل و بهره‌برداری از مخازن سدها به‌عنوان یکی از موارد کلیدی بوده که تلاش‌های زیادی جهت برآورد آن با روش‌های مختلف انجام‌گرفته است.

برآورد میزان رواناب در یک حوزه آبخیز اهمیت زیادی در بهره‌برداری از مخازن و مدیریت آب‌های سطحی دارد و یکی از مهم‌ترین عوامل برآورد و مدیریت ریسک خطرات سیل است. ازجمله ضرورت برآورد میزان رواناب می‌توان به رابطه مستقیم مقدار آب حوزه‌های آبریز با مقدار رواناب آن حوزه نام برد چراکه در طرح‌ها و پروژه‌های آبی علاوه بر ویژگی‌های آن حوزه، به ویژگی‌های رودخانه و هم‌چنین مقدار آبدهی آن و ورودی آن رودخانه نیز نیاز است. پیش‌بینی و مدل‌سازی در مقیاس بلندمدت ریسک و عدم قطعیت بالاتری را با خود به همراه دارد؛ به همین دلیل تحلیل و بررسی این میزان ریسک یکی از عوامل ضروری در مدل‌سازی است که با بالا رفتن دقت، افزایش اعتبار مدل را با خود به همراه دارد (٤١). پیش‌بینی دقیق مقدار دبی جریان در برنامه‌ریزی و مدیریت منابع آب موجود بسیار حائز اهمیت است؛ ازآنجایی‌که سرانه مصرف آب رو به افزایش است، به‌تبع میزان تقاضا نیز به‌سرعت افزایش پیدا می‌کند؛ بنابراین با پیش‌بینی دبی جریان می‌توان به مدیریت بهینه منابع آب پرداخت و به همین علت کسب مدل مناسب با مطمئن‌ترین عملکرد در پیش‌بینی با توجه به شرایط مختلف از اهمیت بالایی برخوردار است (۲۰). پیش‌بینی جریانات سطحی می‌تواند موجب تنظیم جریانات در زمان‌های آتی شود که این امر می‌تواند به مدیریت منابع در دسترس، جهت مصارف مختلف کمک شایانی کند (٤). با توجه به منابع آب شیرین محدود، برآورد دقیق جریان‌های موجود و تغییرات آن از مواردی است که در برنامه‌ریزی منابع آب و تخصیص آب از اهمیت ویژه‌ای برخوردار است (١٤). پیش‌بینی رواناب یکی از مهم‌ترین مسائل در بحث مهندسی آب است که سبب مدیریت، برنامه‌ریزی و جهت تولید انرژی‌های برقابی و تخصیص منابع آب به بخش‌های مختلف می‌گردد. متغیرهای مختلفی بر روی رواناب تأثیرگذار هستند که ازجمله می‌توان به درجه حرارت و مقدار بارندگی اشاره نمود و رابطه بین آن‌ها یک رابطه غیرخطی و پیچیده است (۱۳). درصورتی‌که رواناب برآورد شده و مدل‌سازی انجام‌شده نادرست باشد و خطای چشم‌گیری داشته باشد موجب برنامه‌ریزی و مدیریت نادرست می‌شود و این امر نه‌تنها کمکی به مدیریت منابع آب موجود نمی‌کند بلکه خسارات جبران‌ناپذیر و از دست رفتن این منابع را با خود به دنبال دارد (۲۱). هرچه دقت پیش‌بینی رواناب در یک مدل بالاتر باشد در کنترل عملکرد نیروگاه‌های برقابی و شبکه‌های توزیع برق تأثیر مطلوب‌تری دارد (۹). رواناب یک فرآیند در چرخه هیدرولوژی است که دارای عدم قطعیت بالایی است و یک فرآیند غیرخطی و کاملاً پیچیده است. عوامل متعددی در این فرآیند دخیل هستند که تغییرات زمانی و مکانی این عوامل از دلایل عدم قطعیت در این فرآیند است (۳). به علت پیچیده بودن پدیده‌های فرآیند هیدرولوژی و تأثیراتی

که فعالیت‌های انسانی بر روی این پدیده‌ها می‌گذارد کار مدل‌سازی را به‌علت غیرخطی بودن پدیده سخت‌تر می‌کند (۲۲). پناهی و علیجانی (۲۰۱۳) به پیش‌بینی دبی اوج سیلابی با استفاده از شبکه عصبی مصنوعی و رگرسیون چندمتغیره در حوزه آبخیز مادرسو پرداختند و به این نتیجه رسیدند که در تمامی موارد عملکرد مدل شبکه عصبی مصنوعی بهتر بوده است که این نشان‌دهنده رابطه غیرخطی رواناب است و نتایج مدل شبکه عصبی مصنوعی مورد تأیید و پیشنهاد می‌شود (۲۹). شبیه‌سازی رواناب روزانه حوزه خرم‌آباد با استفاده از مدل‌های شبکه عصبی مصنوعی و سیستم استنتاج فازی عصبی-تطبیقی و رگرسیون چندمتغیره توسط حقی‌زاده و همکاران (۲۰۱۵) انجام شد و پس از بررسی به این نتیجه رسید که مدل استنتاج فازی عصبی-تطبیقی عملکرد و دقت بالاتری نسبت به دو مدل دیگر داشته است (۱۷) کومار و همکاران (۲۰۱٦) در پژوهشی فرآیند رواناب را با استفاده از دو مدل رگرسیون و شبکه عصبی مصنوعی رودخانه بناس هند پرداختند و نتایج بیانگر آن بوده است که مدل شبکه عصبی مصنوعی عملکرد بهتری داشته و نتایج قابل‌قبولی از خود ارائه کرده است (۳۸) در مطالعه ارائه‌شده توسط عدنان و همکاران (۲۰۱۷) به پیش‌بینی جریان ماهانه رودخانه استور پاکستان با استفاده از دو مدل سری زمانی آرما و آریما پرداختند پس از بررسی شاخص‌های خطا دریافتند که نتایج مدل آریما از مدل آرما مطلوب‌تر بوده است (۲). کیا و همکاران (۲۰۱۹) به مدل‌سازی رواناب ناحیه کوهستانی حوزه هراز با استفاده از سیستم استنتاج فازی عصبی-تطبیقی و رگرسیون خطی چندمتغیره پرداختند و به این نتیجه رسیدند که این دو مدل اختلاف معناداری با یکدیگر ندارند؛ ولی در تمام مدل‌ها استنتاج فازی عصبی-تطبیقی نتایج بهتری ارائه کرده است؛ بنابراین استفاده از این مدل پیشنهاد شده است (۲۱). در پژوهشی توسط خزائی و همکاران (۲۰۱۹) به مطالعه تطبیقی چهار مدل شبکه عصبی مصنوعی- نزدیک‌ترین همسایگی- رگرسیون چندمتغیره خطی و سیستم استنتاج فازی عصبی-تطبیقی با تغییر موجک در پیش‌بینی جریان ماهانه پرداختند و به این نتیجه رسیدند که سه مدل مدل شبکه عصبی مصنوعی- نزدیک‌ترین همسایگی و سیستم استنتاج فازی عصبی- تطبیقی نتایج مطلوب‌تری ارائه کرده‌اند. درحالی‌که مدل رگرسیون چندمتغیره خطی به دلیل ایجاد رابطه خطی که بین داده‌های ورودی و خروجی ایجاد می‌کند نمی‌تواند به‌خوبی پیش‌بینی لازم را به‌دقت موردنظر انجام دهد. پس از بررسی نتایج مشخص شد که اضافه کردن داده‌های گذشته جریان، دما و هم‌چنین بارش دقت و عملکرد مدل را به میزان قابل‌توجهی افزایش داده است (۳۰). کانگ و همکاران (۱۹۹۳) با استفاده از دو مدل شبکه عصبی مصنوعی و آرما به پیش‌بینی جریان در بازه‌های زمانی روزانه و ساعتی پرداختند و با تحلیل و بررسی نتایج به این نتیجه رسیدند که مدل شبکه عصبی مصنوعی می‌تواند به‌خوبی به پیش‌بینی جریان بپردازد (۱۹). در پژوهشی توسط وئو و چائو (۲۰۱۰)، جریان ماهانه چند رودخانه در چین را با استفاده از مدل سری زمانی آریما و شبکه عصبی مصنوعی و K نزدیک‌ترین همسایه مدل‌سازی کردند و نتایج بیانگر آن بود که مدل K نزدیک‌ترین همسایه از کارایی مناسب‌تری برخوردار بوده است (٤۲). بشری و وفاخواه (۲۰۱۱) به مقایسه مدل‌های مختلف تحلیل سری‌های زمانی در پیش‌بینی دبی ماهانه حوزه آبخیز کرخه شامل مدل باکس جنکینز، هولت و وینترز و آرما پرداختند و نتایج بیانگر آن بوده است که مدل سری زمانی آرما در دبی‌های بالا و مدل‌های ساده عملکرد بهتری دارد و دلیل آن این است که مدل پایدارتر و نوسان کم‌تری دارد که با نتیجه کار لوو و همکاران (۲۰۰۹) و منتوآ و

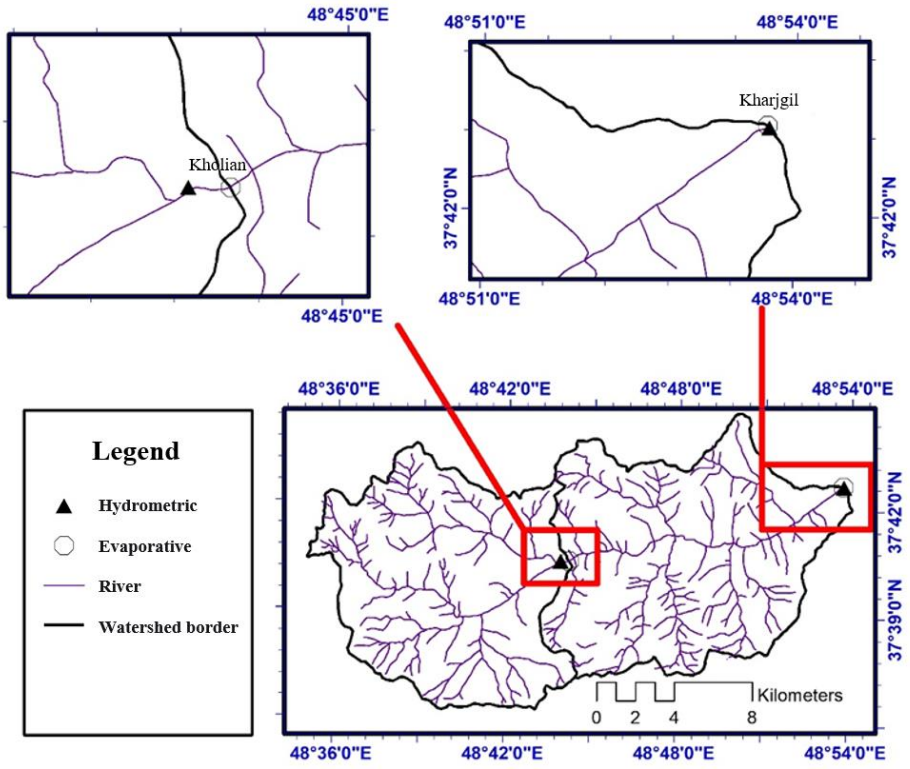
همکاران (۱۹۹۷) مطابقت دارد (۷)(۲۵). مطالعات هادی‌زاده و همکاران (۲۰۱۳) بر روی داده‌های سری زمانی دبی رودخانه گاماسیاب در استان کرمانشاه بیانگر این یافته است که با افزایش مقیاس زمانی از روزانه به فصلی حافظه بلندمدت سری زمانی به شکل قابل‌ملاحظه‌ای کاهش می‌یابد (١٦). در پژوهشی توسط ولی‌پور و همکاران (۲۰۱۳) به مدل‌سازی جریان ماهانه ورودی سد دز با استفاده از مدل خودهمبسته میانگین-متحرک، میانگین متحرک خودهمبسته یکپارچه و شبکه عصبی خودهمبسته پرداختند و نتایج نشان داد عملکرد مدل آریما نسبت به آرما بهتر بوده و به‌طور کل دقت مدل شبکه عصبی مصنوعی از دو مدل دیگر بالاتر بوده است (۳۷) در مطالعه ارائه‌شده توسط وانگ و همکاران (۲۰۱۵) به شبیه‌سازی عددی جریان رواناب سالانه به مخازن سه سد داهوفانگ، بیل یوهی و موپانشا با استفاده از مدل سری زمانی آریما پرداختند و پس از بررسی نتایج کفایت و عملکرد این مدل را تأیید کردند (۳۹). معصوم‌پور و همکاران (۲۰۱۷) به‌کفایت مدل ساریما که یک مدل سری زمانی است برای همه ایستگاه‌های ایران پرداختند و نتایج نشان داد در همه ایستگاه‌های موردمطالعه (به‌جز بوشهر، شهرکرد، بیرجند، امیدیه آغاجاری و رشت) مدل، ساختار همبستگی داده‌ها را به خوبی توصیف و بارش‌های فصلی را در این ایستگاه‌ها به خوبی تعیین می‌کند؛ بنابراین از کفایت لازم برخوردار است (٢٦). در پژوهش ارائه شده توسط میرزاپور و همکاران (۲۰۱۸) به پیش‌بینی دبی ماهانه دو رودخانه نیمه خشک و نیمه مرطوب افرینه کشکان و کاکا رضا پرداختند. در این پژوهش از مدل سری زمانی آریما استفاده شد و نتایج نشان داد که رودخانه افرینه کشکان که در منطقه نیمه خشک قرار دارد؛ پیش‌بینی و نتایج مطلوب‌تری را با استفاده از این مدل داشته‌است (۲۷). در پژوهش ارائه شده توسط صمدی و همکاران (۲۰۲۰) به پیش‌بینی دبی ماهانه ورودی سد بوستان در استان گلستان با استفاده از مدل‌های سری زمانی، شبکه عصبی مصنوعی، ماشین بردار پشتیبان و ۳ مدل ترکیبی انجام شد و نتایج بیانگر آن بود که مدل سری زمانی بهترین مدل و عملکرد را در پیش‌بینی دبی ماهانه از خود نشان داده است (۳۳) در پژوهشی توسط خداخواه و همکاران (۲۰۲۲) به پیش‌بینی سری‌های زمانی دبی ماهانه دو رودخانه پسیخان و خالکایی با استفاده از ٤ مدل سیستم استنتاج فازی عصبی-تطبیقی، کنترل گروهی داده‌ها، سری زمانی ساریما، حداقل مربعات ماشین بردار پشتیبان پرداختند و پس از بررسی نتایج دریافتند که مدل سری زمانی ساریما بهترین عملکرد را در پیش‌بینی دبی جریان در شرایط خشکسالی از خود نشان می‌دهد (۲۰) جندقی و همکاران (۲۰۲۱) با استفاده از دو مدل تابع انتقال و باکس جنکینز به مدلسازی و پیش‌بینی مقادیر دبی دو حوزه رامیان و گالیکش حوزه گرگانرود پرداختند و نتایج بیانگر آن بود که تابع انتقال نسبت به تابع باکس جنکینز عملکرد بهتری داشته است (۱۸) در پژوهش ارائه شده توسط بهارتاچارپا و سولماتین (۲۰۰۵) رابطه سطح آب- دبی رودخانه توسط دو مدل شبکه عصبی مصنوعی و درخت تصمیم M5 مورد بررسی قرار گرفت و نتایج نشان دادند که مدل درخت تصمیم M5 دقت و عملکرد بهتری از خود نشان داده است (۸). قربانی و همکاران (٢٠١٦) در پژوهشی به ارزیابی مدل‌های هیدرولوژیک و داده‌کاوی در شبیه‌سازی و پیش‌بینی جریان ماهانه رودخانه ایستگاه هیدرومتری ارازکوسه پرداختند که در آن سه مدل IHACRES و نزدیک‌ترین همسایگی و درخت تصمیم M5 مورد استفاده قرار گرفت. پس از بررسی به این نتیجه رسیدند که مدل‌های پایه داده‌ای نسبت به مدل‌های هیدرولوژیک دقت و عملکرد بهتری دارند و

نشان‌دهنده آن است که مدل KNN و M5 نتایج قابل‌قبول‌تری نسبت به مدل IHACRES داشته‌است (۱۵). نعیمی و همکاران (۲۰۱٦) در پژوهشی برآورد جریان ماهانه در حوزه‌های فاقد آمار با استفاده از پارامترهای اقلیمی و فیزیوگرافی حوزه استان گلستان با دو مدل درخت تصمیم M5 و رگرسیون چندمتغیره انجام شد و نتایج بیانگر آن بوده‌است که مدل درخت M5 با توجه به سادگی، قابلیت و دقت خوبی در مدلسازی و تخمین آبدهی رودخانه داشته‌است (۲۸) در مطالعه ارائه‌شده توسط سالاری‌جزی و همکاران (۲۰۱٦) پیش‌بینی جریان روزانه حوزه آبخیز گالیکش با استفاده از مدل‌های داده محور برنامه‌ریزی بیان ژن، مدل درخت M5 و مدل ANN صورت گرفت که پس از بررسی به این نتیجه رسیدند که مدل درخت تصمیم M5 به‌دلیل داشتن ضریب همبستگی بالاتر و مقدار خطای کم‌تر از دقت و عملکرد بالاتری برخوردار است. به‌طور کل نتایج شاخص‌های آماری نشان‌دهنده برتری نسبی مدل M5 نسبت به GEP و ANN است و هر سه مدل جریان بیش‌تری نسبت به مقادیر مشاهداتی پیش‌بینی کرده‌اند اما هر ۳ مدل به جهت مدل‌سازی جریان توصیه می‌شود (۳۲) اسمعیلی گیساوندانی و همکاران (۲۰۱۸) در پژوهشی به تحلیل منطقه‌ای سیلاب با مقایسه مدل‌های الگوریتم درخت تصمیم M5 و رگرسیونی پرداختند و نتایج به‌دست آمده نشان‌دهنده آن بود که مدل درخت M5 عملکرد و دقت قابل‌قبولی داشته و در مرحله صحت‌سنجی مطابقت بیش‌تری نسبت به مدل رگرسیونی با دبی سیلاب داشته‌است (۱۲). در پژوهشی توسط عرب و همکاران (۲۰۱۸) به مقایسه دو مدل ناپارامتری KNN و M5 در پیش‌بینی آبدهی دو رودخانه کرج پرداختند. نتایج بیانگر آن بود که مقادیر محاسبه شده همخوانی مناسبی داشته‌اند و هردو مدل توانایی بالایی در برآورد از خود نشان دادند و با مقایسه شاخص‌های آماری برتری نسبی مدل M5 را نسبت به مدل KNN نشان داد که با نتایج پژوهش‌های ستاری و همکاران (۲۰۱۳)، ظهیری و قربانی (۲۰۱۲) همخوانی دارد ولی از آن‌جایی‌که این برتری معنادار نبوده‌است به همین دلیل هر دو مدل جهت برآورد جریان رودخانه مورد استفاده قرار می‌گیرد (۵).

با توجه به سوابق پژوهش بیان شده در هیچ‌یک از پژوهش‌ها سه مدل رگرسیون چندمتغیره خطی، سری زمانی و درخت تصمیم M5 در دو بازه زمانی ماهانه و فصلی مورد مقایسه و بررسی قرار نگرفته و هم‌چنین نتایج مدل‌ها با استفاده از دو رویکرد داده‌های ورودی یک‌متغیره و چندمتغیره مقایسه نشده‌اند. از این‌رو هدف از این پژوهش بررسی نتایج پیش‌بینی رواناب با استفاده دو رویکرد می‌باشد؛ در رویکرد اول در دو بازه زمانی ماهانه و فصلی مقدار رواناب با استفاده از سه مدل رگرسیون چندمتغیره خطی، سری زمانی و درخت تصمیم M5 صورت می‌گیرد؛ متغیرهای ورودی در رویکرد اول بارش، دما و دبی با سه گام تاخیر زمانی می‌باشد و نتایج هر مدل در این رویکرد به‌صورت جداگانه بررسی و مورد تحلیل قرار می‌گیرد. در رویکرد دوم نیز در دو بازه زمانی ماهانه و فصلی مقدار رواناب با استفاده از سه مدل رگرسیون چندمتغیره خطی، سری زمانی و درخت تصمیم M5 صورت می‌گیرد با این تفاوت که در رویکرد دوم متغیر ورودی به مدل تنها مقدار دبی با سه گام تاخیر زمانی می‌باشد و نتایج هر مدل در رویکرد دوم مورد تحلیل و بررسی قرار می‌گیرد در انتها به مقایسه نتایج رویکرد اول و دوم با توجه به بازه‌های زمانی ماهانه و فصلی و هم‌چنین مقایسه نتایج و عملکرد هر مدل در هریک از رویکردها و تأثیر متغیر ورودی به مدل پرداخته می‌شود.

مواد و روش‌ها

منطقه موردمطالعه در این پژوهش حوزه آبخیز ناورود اسالم با مساحت حدود ۳۰۷ کیلومترمربع است که این حوزه در منطقه غرب گیلان در محدوده شهرستان تالش، بین طول‌های جغرافیایی ٤٨ درجه و ٣٦ دقیقه تا ٤٨ درجه و ٥٤ دقیقه شرقی و عرض‌های جغرافیایی ٣٧ درجه و ٣٦ دقیقه تا ٣٧ درجه و ٤٥ دقیقه شمالی قرارگرفته است. این حوزه در دامنه‌های شرقی سلسله جبال البرز واقع‌شده است و از سمت شمال به حوزه آبخیزگرگانرود، از جنوب به حوزه‌های خاله‌سرا و دیناچال، از غرب به حوزه آریاچای و از شرق به دریای خزر متصل شده است. پوشش گیاهی حوزه آبخیزناورود از دو تیپ جنگل و مرتع تشکیل شده است که دارای بارندگی نسبتاً خوبی در طول سال است میانگین بارش سالانه در کل حوزه ٨٣٦ میلی‌متر و متوسط درجه حرارت سالانه ١٠/٧ درجه سانتی‌گراد است (٦). پوشش جنگلی آن از نوع پهن‌برگان مانند گونه‌های بلوط، راش، ممرز، نمدار، شیردار و غیره و پوشش مرتعی آن از نوع شبدر، علف بره، درمنه، گون، بومادران، اسپرس، علف باغی، گل قاصدک و غیره است.



شکل ١- نقشه موقعیت حوزه آبخیز ناورود و زیر حوزه‌های آن.

Figure 1. Location of Guilan province and its sub-basin.

این حوزه در سال ١٣٦٩ پس از بررسی‌هایی توسط کارشناسان، به عنوان معرف منطقه غرب گیلان در نظر گرفته شده، ایستگاه‌های تبخیرسنجی ناو و خرجگیل به‌ترتیب در سال‌های ١٣٦٧ و ١٣٦٨ تاسیس گردید و در سال ١٣٧٦ با احداث ایستگاه‌های هیدرومتری و تبخیرسنجی خلیان واقع در مرکز ثقل

حوزه شبکه هیدروکلیماتولوژی حوزه معرف تکمیل گردید. داده‌های مربوط به دبی، بارش و دما ماهانه و فصلی ایستگاه‌های خرجگیل و خلیان از آب منطقه‌ای استان گیلان جمع‌آوری و مورد استفاده قرار گرفته است.

مدل سری زمانی: استفاده از مدل سری‌های زمانی بیش از چهار دهه پیش با مدل باکس و جنکینز شروع شده است. مدل سری‌های زمانی در هیدرولوژی توسط مک کرچار و دولر اولین گام مهم خود را طی کرده است که مدل ساریما و پارما را برای شبیه‌سازی فصلی رودخانه ارائه کردند (۲٤). داده‌های سری زمانی از هم مستقل نیستند و در مشاهدات همبستگی وجود دارد که این امر سبب شد که در پژوهش‌های مربوط به پیش‌بینی مورداستفاده قرار بگیرد (۱). در سری‌های زمانی فرض بر این است که رفتار و الگوی داده‌های موجود در آینده نیز به همان صورت است و با این فرض به پیش‌بینی سری‌های زمانی آینده می‌پردازد. در سری‌های زمانی دو هدف اساسی مدنظر است، هدف اول مدل کردن مکانیسم تصادفی که سری زمانی را تشکیل داده و هدف دوم پیش‌بینی مقادیر آینده است که با استفاده از اطلاعات گذشته صورت می‌گیرد (۱۰).

در تحلیل یک سری زمانی، تغییرات به ٤ عامل بستگی دارد؛ عامل اول روند است که این تغییرات اصولاً به‌صورت صعودی و یا نزولی است و شامل تغییرات درازمدت در میانگین سری زمانی است، عامل دوم نیز تغییرات فصلی است که این تغییرات در کوتاه‌مدت و دوره‌های کم‌تر از یک سال اتفاق می‌افتد و شامل تغییراتی است که به‌صورت منظم و چرخه‌ای در دوره‌های متناوب رخ می‌دهد، عامل سوم شامل تغییرات دوره‌ای است که این تغییرات برای سری‌های زمانی بیش‌تر از یک سال انجام می‌شود و به‌صورت نوسانی هستند و درنهایت عامل چهارم تغییرات نامنظم است، این تغییرات با الگوی منظم و قانون خاصی رخ نمی‌دهند و تصادفی هستند و زمان وقوع آن غیرقابل‌پیش‌بینی و نامنظم است و نتیجه عوامل تصادفی می‌باشند که به تغییرات نامنظم و با تصادفی یاد می‌شوند (۱).

مدل آریما می‌تواند به دودسته اختیاری غیرفصلی و فصلی تقسیم می‌شود:

۱- آریما غیرفصلی: در این مدل (p,d,q) داریم که در آن q و p به‌ترتیب پارامتر خودهم‌بسته و میانگین متحرک غیرفصلی و d پارامتر تفاضلی کردن جهت ایستاسازی سری زمانی است. شکل رابطه آریما غیرفصلی مطابق رابطه ۱ به‌صورت زیر است:

$$\emptyset(B)Z_t = \emptyset(B)(1-B)Z_t = \theta(B)at \qquad (۱)$$

که در آن، Z_t سری‌های مشاهده شده، $\emptyset(B)$ رتبه چند جمله‌ای p، $\theta(B)$ رتبه چند جمله‌ای q.

۲- آریما فصلی ضربی: در این مدل $(P,D,Q)\times(p,d,q)$ داریم که در آن P و Q پارامترهای خودهم‌بسته و میانگین متحرک فصلی و D پارامتر تفاضلی جهت ایستاسازی سری زمانی است. شکل رابطه آریما فصلی ضربی مطابق رابطه ۲ به‌صورت زیر است:

$$\emptyset_p(B)\emptyset_p(B^s)\Delta^d\Delta_s^D(Z+-\bar{Z}) = \qquad (۲)$$
$$\theta_q(B)\theta_Q(B^s)a_t$$

که در آن، θ_q و $\emptyset_P$ چندجمله‌ای‌های فصلی Q و P هستند.

مدل رگرسیون چندمتغیره خطی: طی پیش‌بینی‌های صورت گرفته مدل رگرسیون چندمتغیره MLR یکی از دسته مدل‌های داده محور است که این مدل‌سازی‌ها برای پیش‌بینی بلندمدت را به‌خوبی انجام می‌دهد

(۲۳). مدل رگرسیون چندمتغیره با ایجاد رابطه خطی بین داده‌های ورودی و خروجی ساختار مدل تشکیل می‌شود که یکی از معایب این مدل ایجاد ارتباط خطی است درصورتی‌که به دلیل این‌که عوامل مختلفی در پیش‌بینی دخیل هستند رابطه یک رابطه غیرخطی است (۳۵).

در مطالعات و مدل‌سازی آماری یک عامل با عاملی دیگر همبستگی گرفته می‌شود اگر بخواهیم عوامل دیگری را هم تأثیر بدهیم و از چند عامل استفاده کنیم استفاده از مدل رگرسیون چندمتغیره بهترین رویکرد است (۲۸). که معادله این مدل مطابق رابطه ۳ به‌صورت زیر است:

$$y=b_0+b_1 x_1+b_2 x_2+\cdots+b_m x_m \quad (۳)$$

مدل درخت تصمیم M5: مدل درخت تصمیم M5 نخستین بار توسط کوینلان (۱۹۹۲) ارائه شد و این مدل کاربرد زیادی در مسائل مهندسی بخصوص در مدل‌سازی و پیش‌بینی رواناب دارد (۳٤). این مدل روشی است که مسائل غیرخطی و پیچیده را به‌راحتی مدل می‌کند و ازآن‌جایی‌که پدیده‌های مربوط به علوم آب، پدیده‌هایی غیرخطی هستند این مدل عملکرد قابل قبولی در این زمینه خواهد داشت (۳٦).

مدل درخت تصمیم M5 مدل چندمرحله‌ای است که داده‌های ورودی را به دسته‌های فرعی تقسیم می‌کند؛ این تقسیم‌بندی تا زمانی ادامه پیدا می‌کند که داده‌های ورودی دیگر قابل‌تقسیم به دسته‌های فرعی نباشند. این مدل با استفاده از ویژگی‌های معلوم که توسط مدل تقسیم شد پیش‌بینی موردنظر را انجام می‌دهد (۳٦). برخلاف مدل‌های درخت معمول، مدل درخت تصمیم M5 در هر گره از مدل یک مدل خطی چندمتغیره را برای داده‌های ورودی تشکیل می‌دهد و شامل دو مرحله ایجاد ساختار مدل و هرس مدل است (٤۰).

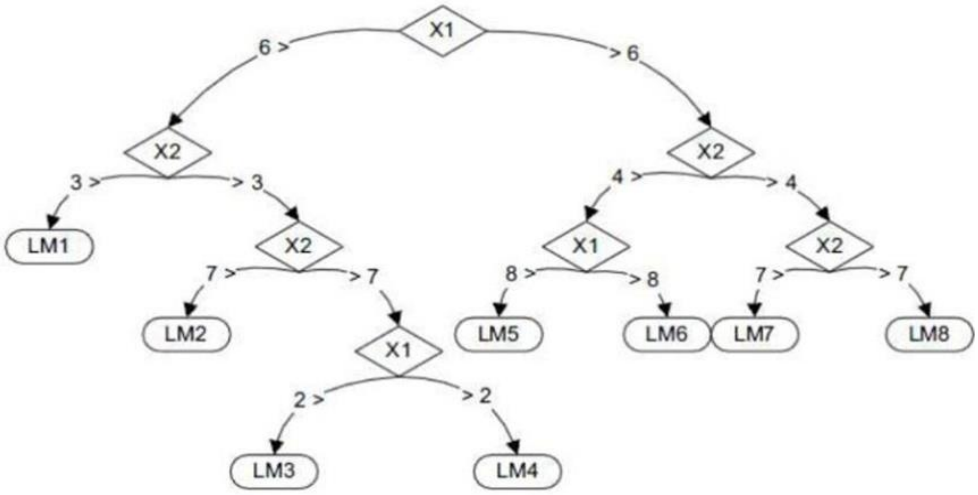
روش کار مدل درخت تصمیم به این صورت است که جهت مدل کردن یک مسأله غیرخطی با تقسیم آن به چند زیر مسأله و ساده‌سازی ساختار مسأله، هر زیر مسأله به‌صورت جداگانه تحلیل و بررسی می‌شود و نتایج هر کدام باهم ترکیب‌شده و به جواب موردنظر دست پیدا می‌کند. پس از ساخت مدل درخت تصمیم جهت عملکرد بهتر مدل لازم است درخت تصمیم مدل شده هرس شود بنابراین جهت بررسی روند هرس، اگر خطای تخمین زده در هر زیر درخت در قسمت ریشه کم‌تر یا برابر میانگین خطای درخت باشد؛ زیر درخت مربوطه هرس می‌شود (۱۲). در این مدل هر گروه نشان‌دهنده ویژگی مشخصی است و در رسم درخت ریشه در ابتدای مدل قرار می‌گیرد (۱۱). مدل درخت تصمیم براساس شاخص انحراف معیار تقسیم می‌شود و یکی از شاخص‌های خطا است. کاهش مورد انتظار هر خطا در گره ارزیابی و محاسبه می‌شود که به‌صورت رابطه ٤ قابل‌محاسبه است (۵).

$$SDR = sd(T) - \sum \frac{|T_i|}{|T|} sd(T_i) \quad (٤)$$

$$sd(T) = \sqrt{\frac{1}{N}\left(\sum_{i=1}^{N} y_i^2 - \frac{1}{N}\left(\sum_{i=1}^{N} y_i\right)^2\right)}$$

که در آن، T نمونه‌های ورودی گره، Sd انحراف معیار، y_i مقدار عددی ویژگی‌های نمونه i، N شماره داده.

مدل درخت تصمیم M5 جهت محاسبه رسوب‌گذاری در رودخانه، مدل‌سازی انتقال رسوب، طبقه‌بندی کاربری اراضی، محاسبه تبخیروتعرق و دیگر موارد مهندسی مورداستفاده قرار می‌گیرد (٤۳).



شکل ۲- مدل درخت تصمیم M5.

Figure 2. Decision tree M5.

جهت انجام این پژوهش دو سری داده مورداستفاده قرار گرفته‌شده است. سری اول داده‌ها شامل داده‌های دبی، بارش و دما به همراه سه‌گام تأخیر هر متغیر و سری دوم داده‌ها تنها شامل دبی تا سه‌گام تأخیر است. با استفاده از سه مدل رگرسیون چندمتغیره، سری زمانی و درخت تصمیم M5 پیش‌بینی دبی در بازه‌های زمانی ماهانه و فصلی برای هر سری از داده‌ها صورت گرفته‌شده است. لازم به ذکر است که ۷۰ درصد داده‌های یادشده به‌عنوان داده‌های آموزش و ۳۰ درصد از داده‌ها نیز به‌عنوان داده‌های صحت‌سنجی و یا آزمون مورداستفاده قرار گرفته‌اند. جهت پیش‌بینی دو مدل رگرسیون چندمتغیره و سری زمانی از نرم‌افزار Minitab 14 و جهت پیش‌بینی مدل درخت تصمیم M5 نرم‌افزار Rapid Miner مورداستفاده قرارگرفته شده است.

شاخص‌های ارزیابی خطا: پس از محاسبات و به‌دست آوردن مقادیر پیش‌بینی با استفاده از ۳ پارامتر زیر مقادیر مشاهداتی و مقادیر محاسباتی مورد مقایسه قرار می‌گیرند تا با استفاده از آن مقادیر خطا در محاسبات موردبررسی قرار بگیرد که این شاخص‌های ارزیابی شامل موارد زیر است:

۱- میانگین انحراف خطا

$$\mathrm{BIAS} = \frac{1}{\mathrm{N}}\sum_{\mathrm{i}=1}^{\mathrm{N}}(\widehat{\mathrm{X}_\imath} - \mathrm{X_i})$$

۲- ضریب همبستگی نش-ساتکلیف

$$\mathrm{NSE} = 1 - \sum_{\mathrm{i}=1}^{\mathrm{N}}(\mathrm{X_i} - \widehat{\mathrm{X}_\imath})^2 / \sum_{\mathrm{i}=1}^{\mathrm{N}}(\mathrm{X_i} - \overline{\mathrm{X}})^2$$

۳- ضریب تعیین

$$R^2 = \frac{\left(n\sum_{\mathrm{i}=1}^{\mathrm{N}}(\mathrm{X_i}\widehat{\mathrm{X}}) - (\sum \mathrm{X_i})(\sum \widehat{\mathrm{X}})\right)^2}{n\sum(\mathrm{X_i})^2\,(\sum \mathrm{X_i})^2 - n\sum(\widehat{\mathrm{X}})^2\,(\sum \widehat{\mathrm{X}})^2}$$

که در آن، $\mathrm{X_i}$ مقدار دبی مشاهداتی، $\widehat{\mathrm{X}_\imath}$ دبی شبیه‌سازی شده، $\overline{\mathrm{X}}$ میانگین دبی مشاهداتی.

نتایج و بحث

در این پژوهش، پیش‌بینی میزان رواناب دو ایستگاه خرجگیل و خلیان موجود در حوزه آبخیز ناورود با استفاده از سه مدل رگرسیون چندمتغیره

خطی، سری زمانی و مدل درخت تصمیم M5 در مقیاس زمانی ماهانه و فصلی با استفاده از دو رویکرد یادشده (رویکرد اول پیش‌بینی با استفاده از متغیرهای دبی، بارش و دما با سه‌گام تأخیر و رویکرد دوم دبی با سه‌گام تأخیر) صورت گرفته‌شده است و نتایج حاصله از هر مدل به‌صورت خلاصه در جدول‌های ۱ تا ٤ ارائه گردید.

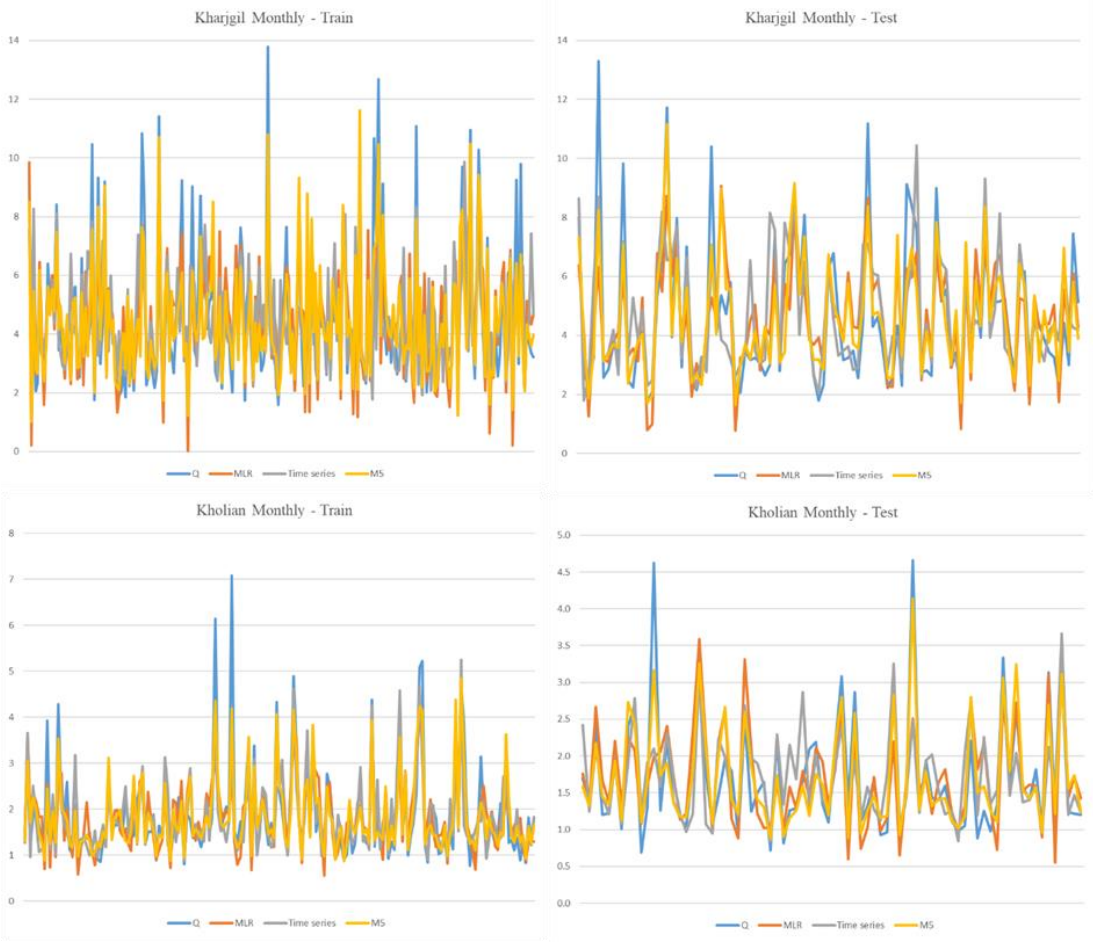
نتایج حاصل از مدل‌سازی رویکرد اول داده‌ها در بازه زمانی ماهانه در جدول ۱ و شکل ۳ به نمایش درآمده که شامل متغیرهای ورودی دبی، بارش و دما تا سه‌گام تأخیر بودند همان‌طور که در جدول ۱ نشان‌دهنده شده نتایج بیانگر این است که مدل درخت تصمیم در هر دو ایستگاه خرجگیل و خلیان، هم در داده‌های آموزش و هم داده‌های آزمون از عملکرد بالاتری نسبت به دو مدل دیگر برخوردار است و این امر بیانگر توانمندی مدل درخت تصمیم M5 در مقایسه با نتایج دو مدل رگرسیون چندمتغیره خطی و سری زمانی است. در بررسی نتایج داده‌های مدل رگرسیون چندمتغیره خطی و سری زمانی در بازه‌های زمانی ماهانه با استفاده از داده‌های سری اول نتایج بیانگر آن بوده که هر دو مدل، عملکرد مشابهی را از خود نشان می‌دهند این در حالی است که در برخی موارد مدل رگرسیون چندمتغیره و در برخی موارد دیگر مدل سری زمانی از خود برتری نشان می‌دهند و هیچ‌یک از این دو مدل برتری مطلقی نسبت به دیگری نداشته است ازاین‌رو جهت پیش‌بینی دبی با توجه به نتایج بیان‌شده در رویکرد اول از بین سه مدل مقایسه شده مدل درخت تصمیم M5 توصیه می‌شود.

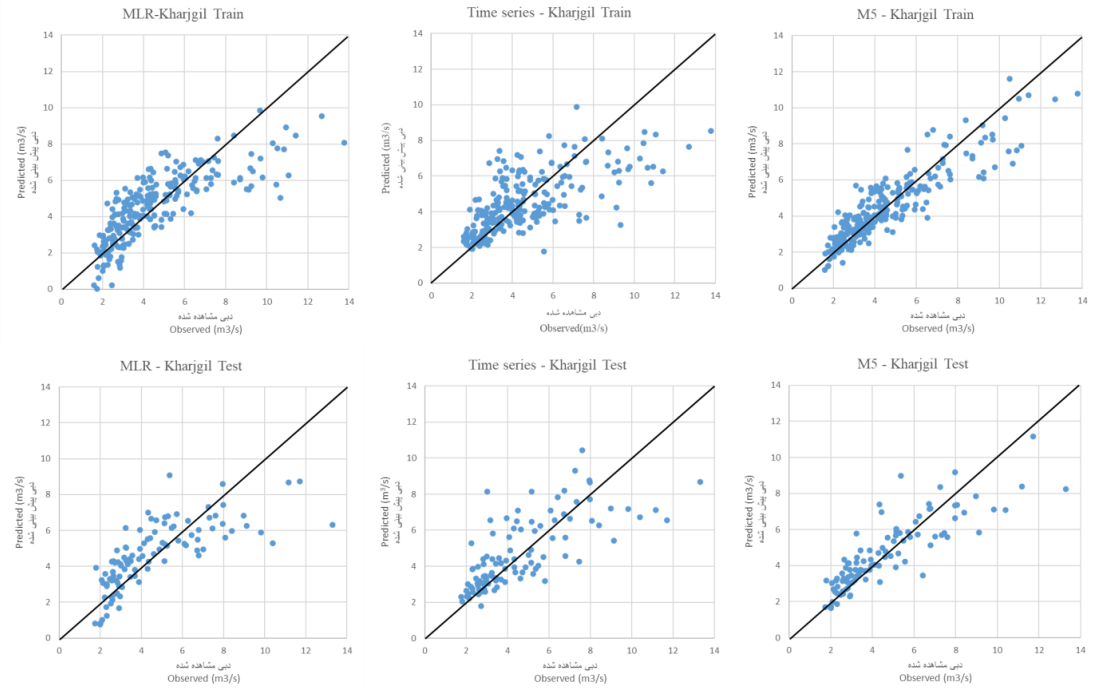
جدول ۱- نتایج ارزیابی مدل‌های رگرسیون چندمتغیره خطی، سری زمانی و درخت تصمیم M5 در پیش‌بینی دبی ماهانه در رویکرد اول.

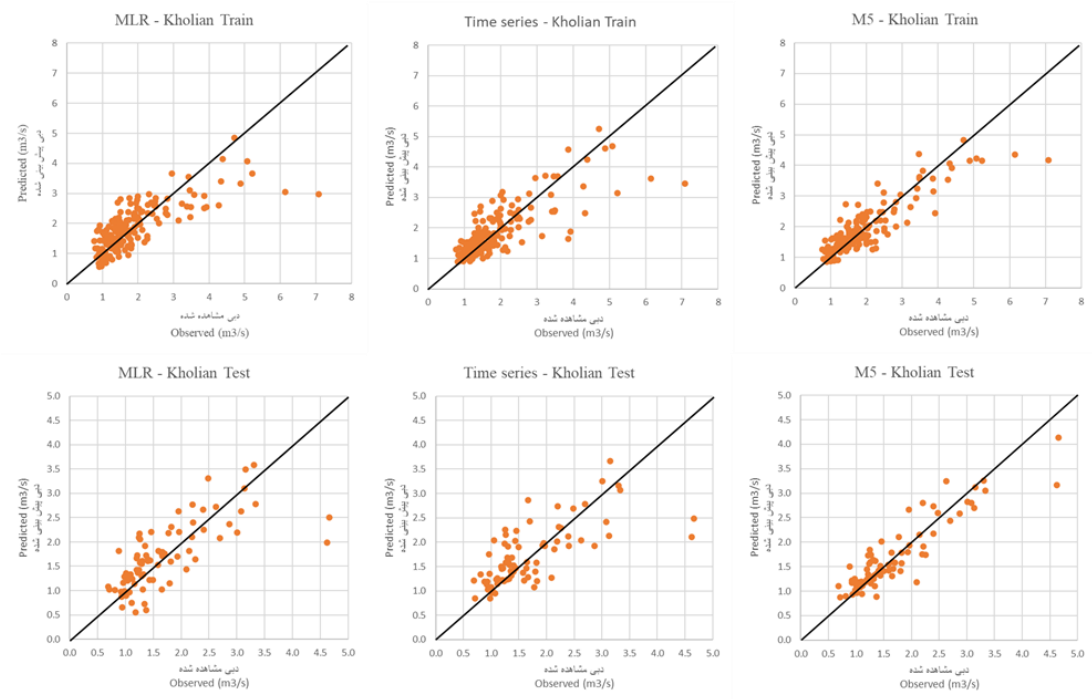
Table 1. The results of Multivariate linear regression, Time series and decision tree M5 methods to simulate the monthly discharge in the first approach.

ایستگاه Station	ماهانه خرجگیل Kharjgil Monthly						ماهانه خلیان Kholian Monthly					
شاخص خطا	آموزش Train			آزمون Test			آموزش Train			آزمون Test		
	R^2	NSE	MBE	R^2	NSE	MBE	R^2	NSE	MBE	R^2	NSE	MBE
MLR	0.62	0.62	-0.01	0.55	0.55	0.03	0.58	0.58	0.00	0.52	0.51	0.01
Time Series	0.47	0.47	-0.05	0.53	0.52	0.10	0.63	0.63	-0.02	0.49	0.49	0.02
M5	0.81	0.80	-0.04	0.72	0.72	0.01	0.80	0.79	-0.01	0.82	0.82	0.00

با ارزیابی شاخص میانگین انحراف خطا در جدول ۱ دو ایستگاه خرجگیل و خلیان برای هر سه مدل رگرسیون چندمتغیره خطی، سری زمانی، درخت تصمیم M5 در بازه زمانی ماهانه داده‌های مربوط به آموزش به مقدار بسیار اندکی کم‌برآوردی همراه است و مقدار آن منفی است این در حالی است که این شاخص برای داده‌های مربوط به صحت‌سنجی و یا آزمون دارای مقدار مثبت بوده است که بیش برآوردی این بخش از داده‌ها را به مقدار اندکی شامل می‌شود.


Kharjgil Monthly - Train
Kharjgil Monthly - Test
Kholian Monthly - Train
Kholian Monthly - Test
Q
MLR
Time series
M5


MLR-Kharjgil Train
Time series - Kharjgil Train
M5 - Kharjgil Train
MLR - Kharjgil Test
Time series - Kharjgil Test
M5 - Kharjgil Test
Predicted (m3/s)
Observed (m3/s)



شکل ۳- رابطه بین دبی مشاهداتی و پیش‌بینی شده ماهانه برای رویکرد اول.

Figure 3. The relationship between observed and predicted monthly discharge for first approach.

در پژوهش‌های مشابه می‌توان به بهارتاچاریا و سولماتین (۲۰۰۵) در بررسی رابطه سطح آب و دبی رودخانه با استفاده از مدل شبکه عصبی مصنوعی و درخت تصمیم M5 (۸)، ظهیری و قربانی (۲۰۱۳) در شبیه‌سازی دبی جریان در مقاطع مرکب (٤٣)، نعیمی و همکاران (۲۰۱٦) در برآورد جریان ماهانه در حوزه‌های فاقد آمار با استفاده از مدل رگرسیون چندمتغیره و مدل درخت تصمیم M5 (۲۸)، اسمعیلی گیساوندانی و همکاران (۲۰۱۸) در تحلیل منطقه‌ای سیلاب با استفاده از مدل درخت تصمیم M5 و مدل رگرسیونی، در پیش‌بینی آبدهی رودخانه با استفاده از مدل K نزدیک‌ترین همسایه و مدل درخت تصمیم M5 (۱۲)، اشاره کرد که در تمامی پژوهش‌های نامبرده نتایج حاصله بیانگر برتری مدل درخت تصمیم M5 در مقایسه با سایر مدل‌های مورداستفاده است که مطابق با نتیجه حاصله می‌باشد.

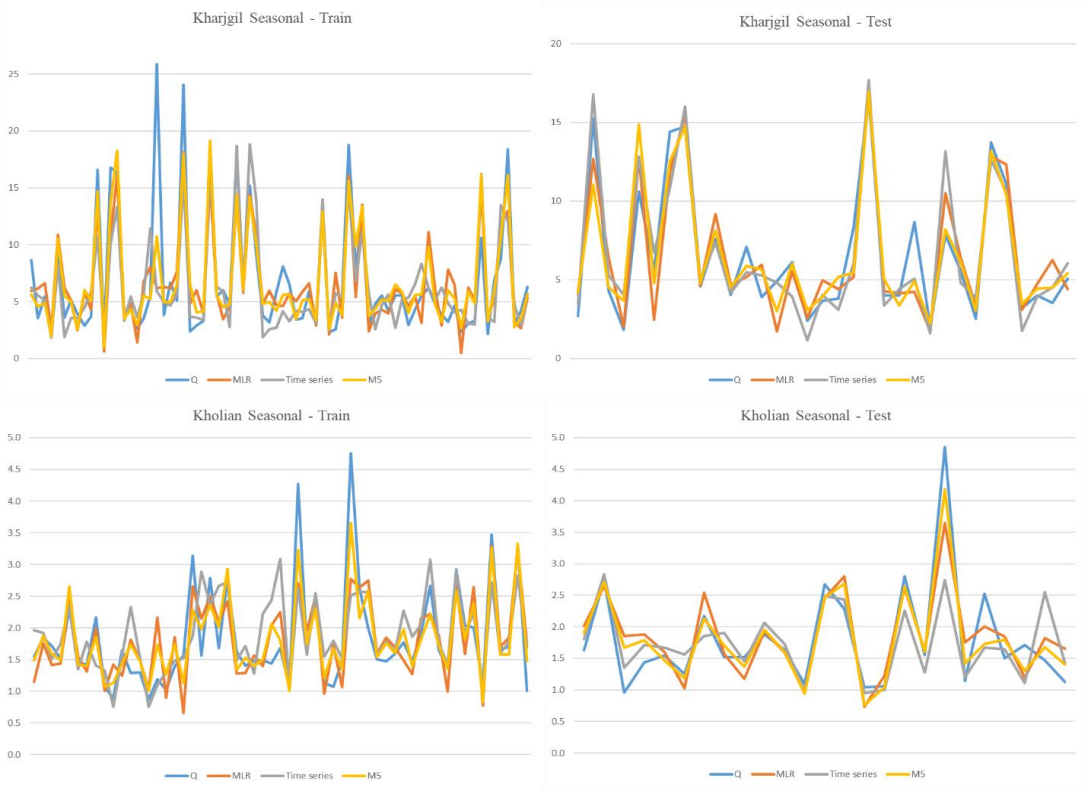
نتایج حاصل از مدل‌سازی رویکرد اول داده‌ها در بازه فصلی ماهانه در جدول ۲ و شکل ٤ به نمایش در آمده؛ با توجه به نتایج درج‌شده در جدول ۲ در داده‌های رویکرد اول در بازه زمانی فصلی مدل درخت تصمیم M5 عملکرد قابل‌قبولی را بین مدل‌های استفاده‌شده از خود نشان داده است این در حالی است که مدل رگرسیون چندمتغیره خطی و مدل سری زمانی نتایج تقریباً مشابهی داشتند ولی به‌طور کل مدل سری زمانی خطای بیش‌تری را نسبت به دو مدل دیگر در پیش‌بینی دبی از خود نشان داده است. درمجموع برای داده‌های مربوط به رویکرد اول مدل درخت تصمیم M5 انتخاب می‌شود که بیانگر توانمندی این مدل در پیش‌بینی فصلی داده‌ها در رویکرد اول است.

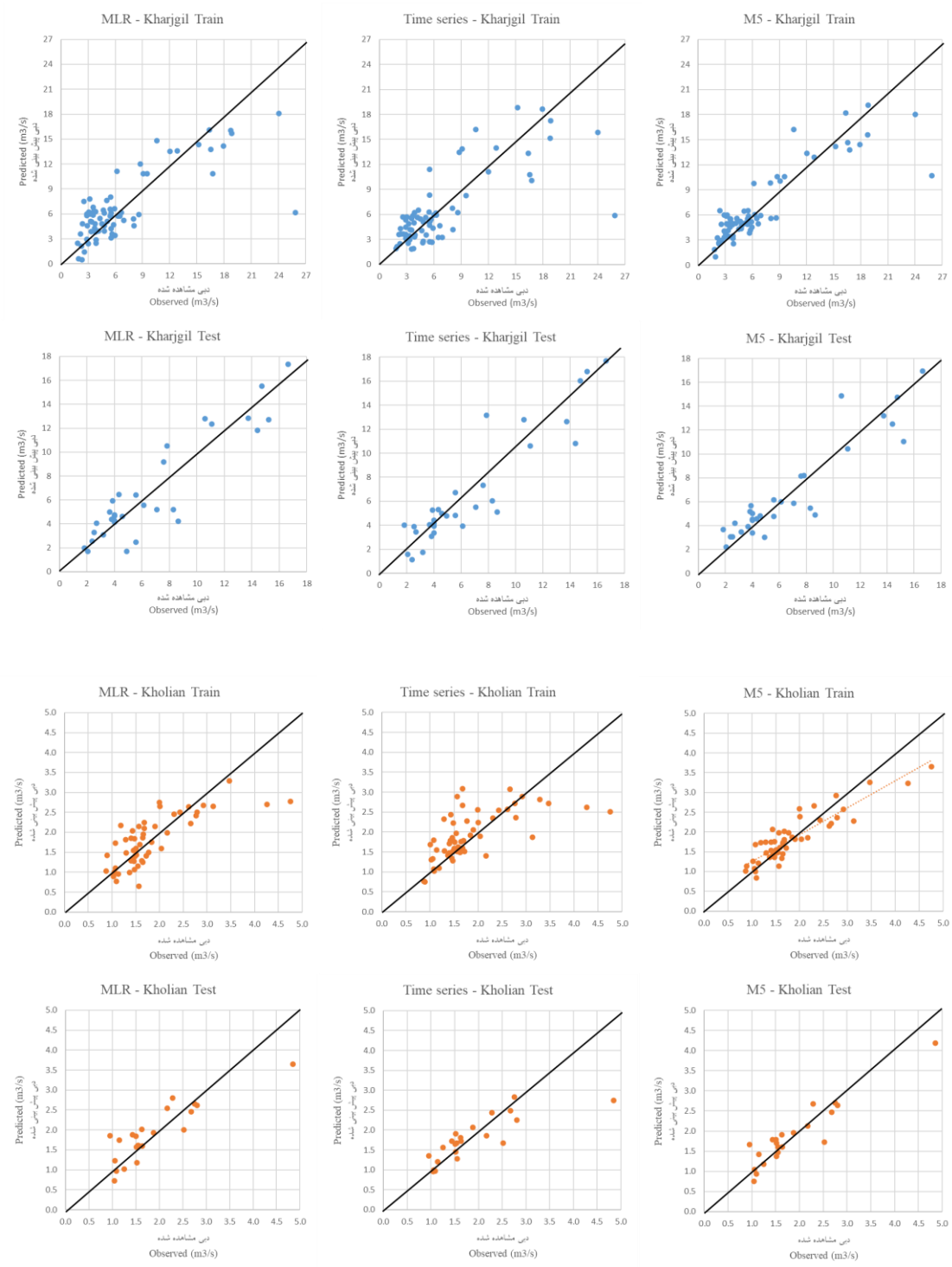
جدول ۲- نتایج ارزیابی مدل‌های رگرسیون چندمتغیره خطی، سری زمانی و درخت تصمیم M5 در پیش‌بینی دبی فصلی در رویکرد اول.

Table 2. The results of Multivariate linear regression, Time series and decision tree M5 methods to simulate the seasonal discharge in the first approach.

ایستگاه Station	فصلی خرجگیل Kharjgil Seasonal						فصلی خلیان Kholian Seasonal					
ضریب خطا	آموزش Train			آزمون Test			آموزش Train			آزمون Test		
	R^2	NSE	MBE	R^2	NSE	MBE	R^2	NSE	MBE	R^2	NSE	MBE
MLR	0.61	0.62	-0.09	0.83	0.82	-0.02	0.57	0.59	-0.02	0.76	0.75	0.03
Time Series	0.57	0.58	-0.42	0.85	0.83	0.06	0.39	0.40	0.09	0.67	0.62	-0.10
M5	0.77	0.78	0.02	0.86	0.86	-0.02	0.80	0.79	-0.01	0.86	0.86	0.00

با توجه به شاخص میانگین انحراف خطا در جدول ۲ برای داده‌های رویکرد اول در بازه زمانی فصلی، زمانی که مقدار این شاخص در داده‌های آموزش مثبت است و بیش‌برآوردی مشاهده‌شده در مقابل مقدار این شاخص در داده‌های آزمون منفی است که با کم‌برآوردی همراه است و تنها برای داده‌های مربوط به مدل رگرسیون چندمتغیره خطی در ایستگاه خرجگیل هم در داده‌های آموزش و هم در داده‌های آزمون کم‌برآوردی حاصل‌شده است. در بررسی نتایج یافته شده در رویکرد اول در هردو بازه زمانی پیش‌بینی‌شده و از بین مدل‌های انتخابی مدل درخت تصمیم می‌تواند نتایج خوبی را از بین مدل‌های استفاده‌شده ارائه دهد که نتیجه یافته شده با پژوهش‌های مشابه مطابقت داشته است و در راستای تائید پژوهش‌ها است.





شکل ٤- رابطه بین دبی فصلی مشاهداتی و پیش‌بینی‌شده برای رویکرد اول.

Figure 4. The relationship between observed and predicted seasonal discharge for first approach.

نتایج حاصل از مدل‌سازی رویکرد دوم داده‌ها در بازه زمانی ماهانه در جدول ۳ و شکل ۵ به نمایش درآمده؛ در داده‌های رویکرد دوم با حذف متغیرهای بارش و دما و تنها با استفاده از ۳ گام تأخیر دبی، پیش‌بینی با استفاده از سه مدل رگرسیون چندمتغیره خطی، سری زمانی و درخت تصمیم M5 صورت گرفته‌شده است و نتایج حاصله از ۳ مدل یادشده به‌صورت مختصر در جدول ۳ برای بازه زمانی ماهانه مربوط به دو ایستگاه خرجگیل و خلیان حوزه آبخیزناورود موردبررسی و مقایسه قرار گرفته‌شده است. بررسی نتایج بیانگر آن بوده که هنگام پیش‌بینی با رویکرد دوم در بازه زمانی ماهانه و تنها با یک متغیر مدل سری زمانی عملکرد بهتری نسبت به دو مدل رگرسیون چندمتغیره خطی و درخت تصمیم M5 داشته است که نتیجه یافته شده در نقطه مقابل داده‌های رویکرد اول قرار دارد این امر نشان‌دهنده عملکرد بالا در مدل سری زمانی هنگام استفاده از یک متغیر در مقایسه با دو مدل دیگر است. با مقایسه نتایج دو مدل رگرسیون چندمتغیره و درخت تصمیم M5 این نتیجه حاصل شد که در هردو ایستگاه، این دو مدل عملکرد تقریباً مشابهی را از خود نشان می‌دهد و هیچ‌یک از این دو مدل نسبت به دیگری برتری مطلقی نداشته است و هر دو مدل عملکرد غیرقابل‌قبولی از خود نشان داده‌اند درنتیجه برای پیش‌بینی با داده‌های رویکرد دوم از میان ۳ مدل بررسی‌شده مدل سری زمانی انتخاب می‌شود.

جدول ۳- نتایج ارزیابی مدل‌های رگرسیون چندمتغیره خطی، سری زمانی و درخت تصمیم M5 در پیش‌بینی دبی ماهانه در رویکرد دوم.

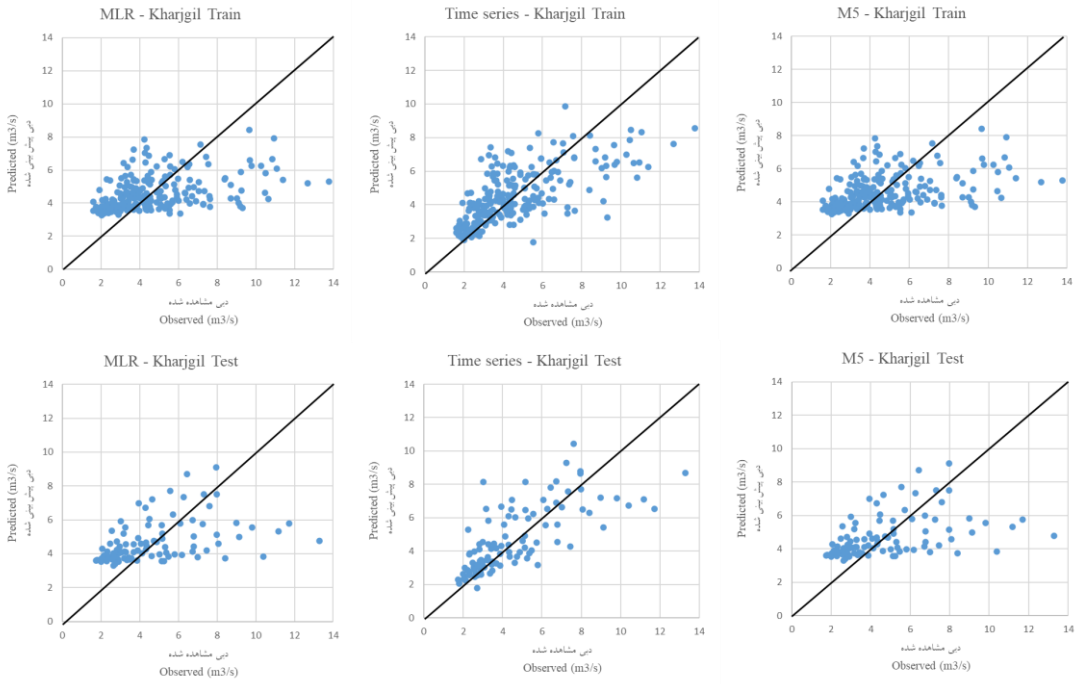
Table 3. The results of Multivariate linear regression, Time series and decision tree M5 methods to simulate the monthly discharge in the second approach.

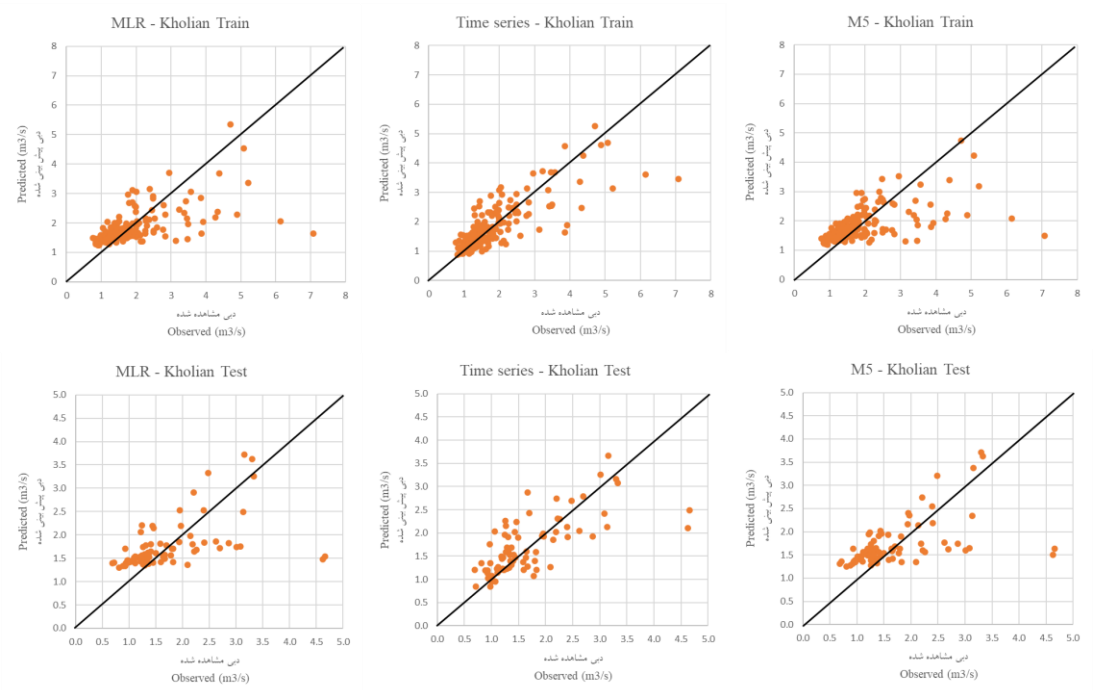
ایستگاه Station	ماهانه خرجگیل Kharjgil Monthly						ماهانه خلیان Kholian Monthly					
ضریب خطا	آموزش Train			آزمون Test			آموزش Train			آزمون Test		
	R^2	NSE	MBE	R^2	NSE	MBE	R^2	NSE	MBE	R^2	NSE	MBE
MLR	0.22	0.22	-0.02	0.20	0.20	0.04	0.36	0.35	-0.02	0.26	0.24	0.05
Time Series	0.47	0.47	-0.05	0.53	0.52	0.10	0.63	0.63	-0.02	0.49	0.49	0.02
M5	0.22	0.22	-0.02	0.20	0.20	0.04	0.32	0.32	-0.01	0.28	0.26	0.03

با توجه به شاخص میانگین انحراف خطا در جدول ۳ نتایج حاصله همانند نتایج رویکرد اول می‌باشد به این صورت که در مقادیر آموزش مقدار این شاخص منفی بوده که با کم‌برآوردی همراه است ولی برای مقادیر آزمون این مقدار مثبت است که نشان از بیش‌برآوردی دارد لازم به ذکر است که مقدار آن اندک بوده و برای هر دو ایستگاه خرجگیل و خلیان روند به همین صورت است.

در پژوهش‌های مشابه می‌توان به بشری و وفاخواه (۲۰۱۱) در مقایسه مدل‌های مختلف سری زمانی شامل مدل باکس جنکینز، هولت و وینترز و آرما (۷)، عدنان و همکاران (۲۰۱۷) در پیش‌بینی جریان ماهانه رودخانه استور پاکستان با استفاده از دو مدل آرما و آریما (۲)، میرزاپور و همکاران (۲۰۱۸) در پیش‌بینی دبی ماهانه رودخانه با استفاده از مدل سری زمانی ساریما (۲۷)، وانگ و همکاران (۲۰۱۵) در شبیه‌سازی جریان رواناب سالانه با استفاده از مدل آریما اشاره کرد (۳۹)؛ که تمامی پژوهش‌های انجام شده به کفایت مدل سری زمانی اشاره دارد که مطابق با نتیجه حاصله از این رویکرد می‌باشد.

Kharjgil Monthly - Train
Kharjgil Monthly - Test
Kholian Monthly - Train
Kholian Monthly - Test
Q
MLR
Time series
M5


MLR - Kharjgil Train
Time series - Kharjgil Train
M5 - Kharjgil Train
MLR - Kharjgil Test
Time series - Kharjgil Test
M5 - Kharjgil Test
Predicted (m3/s)
Observed (m3/s)



شکل ٥- رابطه بین دبی مشاهداتی و پیش‌بینی شده ماهانه برای رویکرد دوم.

Figure 5. The relationship between observed and predicted monthly discharge for second approach.

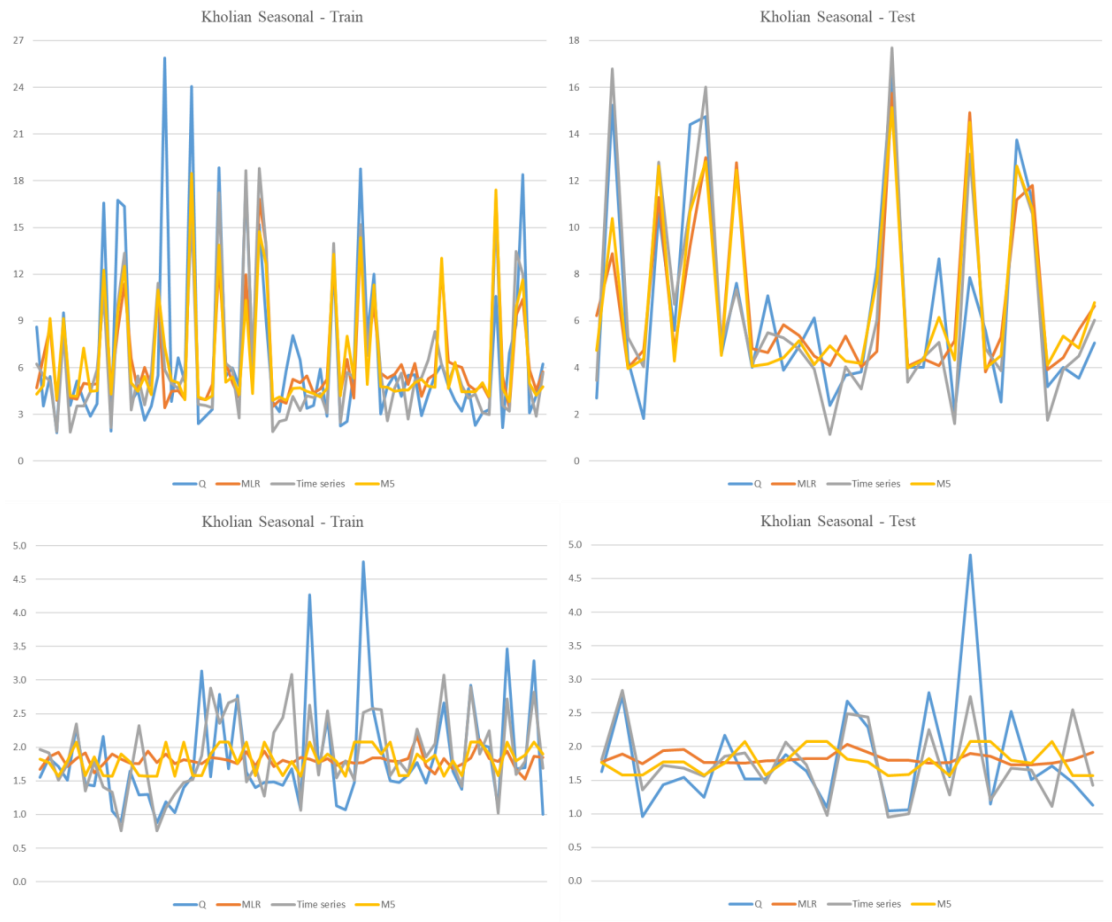
نتایج حاصل از مدل‌سازی رویکرد دوم داده‌ها در بازه زمانی ماهانه در جدول ٤ و شکل ٦ به نمایش در آمده با توجه به نتایج حاصل‌شده داده‌های سری اول در بازه زمانی سه‌ماهه در مدل‌های رگرسیون چندمتغیره خطی و هم‌چنین درخت تصمیم M5 در ایستگاه خلیان عملکرد پایینی از خود نشان داده‌اند که این امر بیانگر این است که این دو مدل برای پیش‌بینی‌های تک‌متغیره مناسب نمی‌باشند و انتخاب نمی‌شوند درحالی‌که مدل سری زمانی عملکرد نسبتاً بهتری در مقابل دو مدل دیگر داشته است و برای رویکرد دوم استفاده از مدل سری زمانی نسبت به سه مدل مورداستفاده شده انتخاب می‌شود.

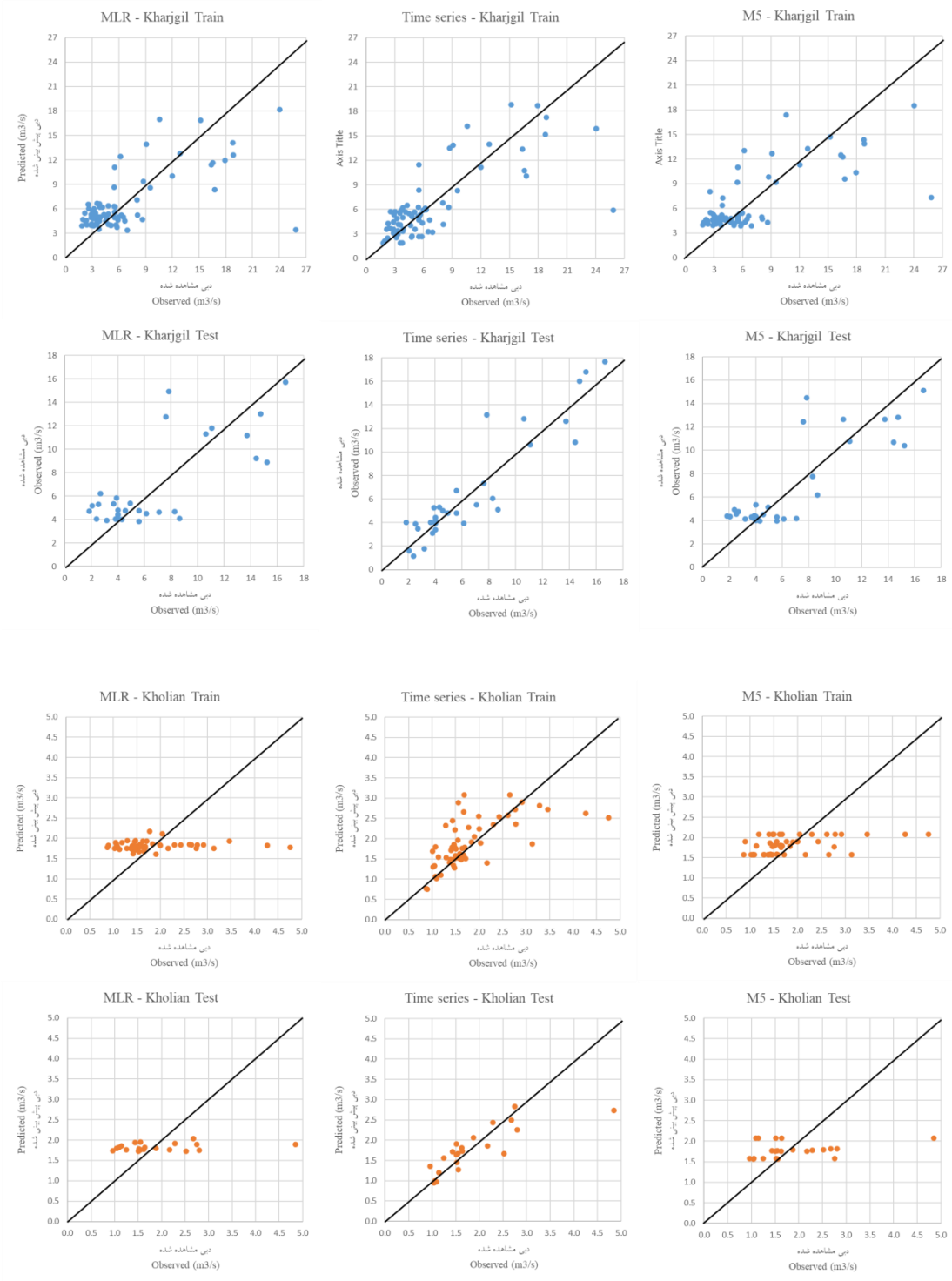
جدول ٤- نتایج ارزیابی مدل‌های رگرسیون چندمتغیره خطی، سری زمانی و درخت تصمیم M5 در پیش‌بینی دبی فصلی در رویکرد دوم.

Table 4. The results of Multivariate linear regression, Time series and decision tree M5 methods to simulate the seasonal discharge in the second approach.

ایستگاه Station	فصلی خرجگیل Kharjgil Monthly						فصلی خلیان Kholian Monthly					
ضریب خطا	آموزش Train			آزمون Test			آموزش Train			آزمون Test		
	R^2	NSE	MBE	R^2	NSE	MBE	R^2	NSE	MBE	R^2	NSE	MBE
MLR	0.45	0.45	-0.10	0.57	0.56	0.20	0.01	0.00	-0.02	0.10	0.05	-0.03
Time Series	0.57	0.58	-0.42	0.85	0.83	0.06	0.39	0.40	0.09	0.67	0.62	-0.10
M5	0.55	0.56	-0.23	0.70	0.70	0.24	0.18	0.15	-0.01	0.08	0.07	-0.07

با توجه به شاخص میانگین انحراف خطا می‌توان نتیجه گرفت که مقدار این ضریب مربوط به ایستگاه خرجگیل برای داده‌های آموزش منفی بوده که با کم‌برآوردی و برای داده‌های آزمون مثبت بوده که با بیش‌برآوردی همراه است این در حالی است که میزان پیش‌بینی برای ایستگاه خلیان هم در داده‌های آموزش و هم داده‌های آزمون منفی (جز در داده‌های آموزش مدل سری زمانی) بوده و نشان از کم‌برآوردی در رویکرد دوم در بازه زمانی فصلی دارد. در پژوهش‌های مشابه، وئو و چائو (۲۰۱۰) به مدل‌سازی جریان ماهانه رودخانه با استفاده از مدل سری زمانی، شبکه عصبی و نزدیک‌ترین همسایه پرداختند که مدل K نزدیک‌ترین همسایه بهترین عملکرد را از خود نشان داده است (٤٢)، هم‌چنین ولی‌پور و همکاران (۲۰۱۳) به مدل‌سازی جریان ورودی سد با مدل سری زمانی و شبکه عصبی نتیجه حاصله نشان‌دهنده کفایت مدل شبکه عصبی مصنوعی بوده (۳۷) که با نتیجه حاصله از رویکرد دوم مغایرت داشته است.





شکل ٦- رابطه بین دبی مشاهداتی و پیش‌بینی‌شده فصلی برای رویکرد دوم.

Figure 6. The relationship between observed and predicted seasonal discharge for second approach.

علت پیش‌بینی با استفاده از دو رویکرد مقایسه نتایج مدل‌های رگرسیون چندمتغیره خطی، مدل سری زمانی، مدل درخت تصمیم M5 در زمان استفاده از یک متغیر ورودی با چند متغیر ورودی در بازه‌های زمانی ماهانه و سه‌ماهه بوده است تا بررسی شود تأثیر متغیرها در دقت و عملکرد هر مدل چه تأثیری دارد.

نتیجه‌گیری کلی

همان‌طور که از جدول نتایج ۱ تا ٤ در بخش نتایج و بحث مشخص است در هنگام پیش‌بینی با استفاده از چند متغیر ورودی که شامل دبی، بارش و دما به همراه سه‌گام تأخیر زمانی است مدل داده‌کاوی درخت تصمیم M5 از میان سه مدل رگرسیون چندمتغیره خطی، سری زمانی و درخت تصمیم M5 بهترین عملکرد و نتیجه را در هر دو بازه زمانی ماهانه و فصلی از خود نشان می‌دهد. اما زمانی که تعداد متغیرهای ورودی به مدل کاهش پیدا می‌کنند و تنها از متغیر دبی به همراه سه‌گام زمانی استفاده می‌شود، مدل درخت تصمیم M5 نتیجه عکس از خود نشان می‌دهد که می‌توان از محدودیت‌های این مدل در تعداد متغیرهای مؤثر به مدل اشاره نمود. در داده‌های سری دوم تنها مدل سری زمانی می‌تواند پیش‌بینی مناسبی از بین سه مدل مورداستفاده ارائه دهد. بنابراین، با توجه به نتایج حاصله مدل درخت تصمیم M5 زمانی برتری دارد که تعداد متغیر ورودی به سیستم بیش‌تر بوده و این امر به این معنا نیست که هرچه تعداد متغیرها بیش‌تر باشد دقت مدل درخت تصمیم M5 بیش‌تر می‌شود؛ منظور از متغیر بیش‌تر، متغیرهایی هستند که بیش‌ترین میزان تأثیر را در نتایج و پیش‌بینی مدل از خود نشان می‌دهند. اما زمانی که تعداد متغیر ورودی به مدل صورت تک‌متغیره بوده مدل درخت تصمیم M5 در هیچ‌یک از بازه‌های زمانی ماهانه و فصلی نتوانسته است قابلیت خوبی را ارائه دهد و حتی نتایج این مدل از مدل سری زمانی پایین‌تر خواهد بود؛ پس در شرایط بیان‌شده و استفاده از تنها یک متغیر با تأخیر زمانی مدل سری زمانی با توجه به نتایج یافته شده توصیه می‌شود.

تقدیر و تشکر

– نویسندگان از داوران این مقاله که با نظرات خود موجب بهبود متن حاضرشده‌اند تشکر می‌نمایند.
– این پژوهش در قالب یک پایان‌نامه در دانشگاه علوم کشاورزی و منابع طبیعی گرگان صورت گرفته است.

داده‌ها، اطلاعات و دسترسی

– داده‌های این پژوهش مربوط به پایان‌نامه فوق‌لیسانس نویسنده اول است که با مکاتبه با نویسنده مسئول قابل‌دسترسی می‌باشند.
– داده‌های این متن مرتبط با پروژه تحقیقاتی با عنوان ارزیابی مدل یادگیرجمعی برای افزایش دقت پیش‌بینی رواناب مبتنی بر سری‌های زمانی و مدل‌های هوشمند در دانشگاه علوم کشاورزی و منابع طبیعی گرگان می‌باشد. دسترسی به داده‌ها و اطلاعات صرفاً برای داوران/دبیرتخصصی/سردبیر نشریه و بنابر درخواست ایشان میسر خواهد بود.

تعارض منافع

در این مقاله تعارض منافعی وجود ندارد و این مسأله مورد تأیید همه نویسندگان است.

مشارکت نویسندگان

نویسنده اول: دسترسی به داده‌ها و داده‌برداری، آماده‌سازی داده‌ها، مشارکت در آنالیزها. نویسنده دوم: طرح پژوهش و روش‌شناسی، مشارکت در آنالیزها. نویسنده سوم: اصلاح و نهایی‌سازی مقاله و مشارکت در آنالیزها. نویسنده چهارم: اصلاح و نهایی‌سازی مقاله و مشارکت در آنالیزها.

اصول اخلاقی

نویسندگان اصول اخلاقی را در انجام و انتشار این اثر عملی رعایت نموده‌اند و این موضوع مورد تأیید همه آن‌ها می‌باشد.

حمایت مالی

– این پژوهش از حمایت مستقیم مالی برخوردار نبوده است، ولی از امکانات دانشگاه علوم کشاورزی و منابع طبیعی گرگان استفاده گردیده است.

منابع

1. Abdolahnezhad, K. 2015. Forecasting of Monthly Sum-raining by Stochastic Models in Time Series. Geographical Planning of Space, 5: 17. 15-25. (In Persian)
2. Adnan, R.M., Yuan, X., Kisi, O., and Curtef, V. 2017. Application of time series models for streamflow forecasting. Civil and Environmental Research, 9: 3. 56-63.
3. Ahmad, S., and Simonovic, S.P. 2005. An artificial neural network model for generating hydrograph from hydro-meteorological parameters. Journal of Hydrology, 315: 1-4. 236-251.
4. Akhtar, M.K., Corzo, G.A., Van Andel, S.J., and Jonoski, A. 2009. River flow forecasting with artificial neural networks using satellite observed precipitation pre-processed with flow length and travel time information: case study of the Ganges river basin. Hydrology and Earth System Sciences, 13: 9. 1607-1618.
5. Arab, S., Khashei, S.A., Pourreza, B.M., and Hashemi, S.R. 2018. Comparison of Two Nonparametric Models, K-nearest neighbor and M5 Decision Tree in Forecasting the River Discharge in the Karaj Catchment. Watershed Management Research Journal. 30: 4. 47-58. (In Persian)
6. Asadi, H., Honarmand, M., Vazifedoust, M., and Mousavi, A. 2017. Assessment of Changes in Soil Erosion Risk Using RUSLE in Navrood Watershed, Iran. J. Agric. Sci. Tech. 19: 231-244.
7. Bashari, M., and Vatankhah, M. 2011. Comparison of different time series analysis methods for forecasting monthly discharge in Karkheh watershed. Irrigation and Water Engineering, 1: 2. 75-86. (In Persian)
8. Bhattacharya, B., and Solomatine, D.P. 2005. Neural networks and M5 model trees in modelling water level–discharge relationship. Neurocomputing, 63: 381-396.
9. Cheng, C.T., Niu, W.J., Feng, Z.K., Shen, J.J., and Chau, K.W. 2015. Daily reservoir runoff forecasting method using artificial neural network based on quantum-behaved particle swarm optimization. Water, 7: 8. 4232-4246.
10. Cryer, J.D. 1992. Time series Analysis. Translated by Niroomand, H.A., Mashhad University Publication, 404p. (In Persian)
11. Eshghi, P., Farzadmehr, J., Dastorani, M.T., and Arabasadi, Z. 2016. The effectiveness of intelligent models in estimating the river suspended sediments (Case Study: Babaaman Basin, Northern Khorasan. JWMR. 2017; 7: 14. 95-88. (In Persian)
12. Esmaeili, H., Akhond Ali, A.M., Zarei, H., and Taghian, M. 2018. Regional Flood Analysis Via Comparison of The M5 Decision Tree Algorithm and Regression Models. Irrigation Sciences and Engineering, 40: 4. 183-195. (In Persian)
13. Fathi, P., Mohammadi, Y., and Homaei, M. 2009. Intelligent modeling of monthly flow time series into vahdat dam in sanandaj city. Journal of Water and soil, 21: 1. 209-220. (In Persian)
14. Ghorbani, K., Naeimi Kalourazi, Z., Salarijazi, M., and Dehghani, A.A. 2016. Estimation of monthly discharge using climatological and physiographic parametesr of ungauged basin. Journal of Water and Soil Conservation, 23: 3. 207-224. (In Persian)
15. 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 Persian)
16. Hadizadeh, R., Eslamian, S., and Chinipardaz, R., 2013, Investigation of long-memory properties in streamflow time series in Gamasiab River, Iran', Int. J. Hydrology Science and Technology, 3: 4. 319-350.
17. Haghizadeh, A., Mohammadlou, M., and Noori, F. 2015. Simulation of rainfall-runoff process using multilayer perceptron and adaptive neuro-fuzzy interface system and multiple regression (Case Study: Khorramabd Watershed. Iranian journal of Ecohydrology, 2: 2. 233-243. (In Persian)
18. Jandaghi, N., Azimmohseni, M., and Ghareh Mahmoodlu, M. 2021. Rainfall-runoff process modeling using time series transfer function. Environmental Erosion Research Journal, 11: 2. 111-128. (In Persian)
19. Kang, K.W., Kim, J.H., Park, C.Y., and Ham, K.J. 1993. Evaluation of hydrologic forecasting system based on neural network model. In Proceedings of the Congress-international Association for research. 1: 257-257.
20. Khodakhah, H., Aghelpour, P., and Hamedi, Z. 2022. Comparing linear and non-linear data-driven approaches in monthly river flow prediction, based on the models SARIMA, LSSVM, ANFIS, and GMDH. Environmental Science and Pollution Research, 29: 15. 21935-21954.
21. Kia, I., Emadi, A.R., and Gholami, M. 2019. Rainfall-Runoff Modeling by Adaptive Neuro-Fuzzy Inference System (ANFIS) and Multi-Variable Linear Regression (MLR). Irrigation and Water Engineering, 9: 4. 39-51. (In Persian)
22. Kratzert, F., Klotz, D., Brenner, C., Schulz, K., and Herrnegger, M. 2018. Rainfall–runoff modelling using long short-term memory (LSTM) networks. Hydrology and Earth System Sciences, 22: 11. 6005-6022.
23. Krstanovic, P.F., and Singh, V.P. 1991. A univariate model for long-term streamflow forecasting. Stochastic hydrology and hydraulics, 5: 3. 189-205.
24. Laux, P., Vogl, S., Qiu, W., Knoche, H. R., and Kunstmann, H. 2011. Copula-based statistical refinement of precipitation in RCM simulations over complex terrain. Hydrology and Earth System Sciences, 15: 7. 2401-2419.
25. Mantua, N.J., Hare, S.R., Zhang, Y., Wallace, J.M., and Francis, R.C. 1997. A Pacific interdecadal climate oscillation with impacts on salmon production. Bulletin of the american Meteorological Society, 78: 6. 1069-1080.
26. Masoumpour Samakosh, J., Jalilian, A., and Yari, E. 2017. The analysis of seasonal precipitation time series in Iran. Physical Geography Research Quarterly, 49: 3. 457-475. (In Persian)
27. Mirzapour, H., Haghizadeh, A., and Alijani, R. 2018. The Evaluation of the Performance of SARIMA Time Series Model in the Simulation of the Average Monthly Discharge of the Kashkan Afrineh and Kakareza Rivers (Lorestan province). Hydrogeomorphology, 4: 55. 153-169. (In Persian)
28. Naeimi Kalourazi, Z., Kh, G., Salarijazi, M., and Dehghani, A.A. 2016. Estimation of monthly discharge using climatic and physiographic parameters of ungauged basins. Journal of Water and Soil Conservation, 23: 3. 207-224. (In Persian)
29. Panahi, A., and AIijani, B. 2013. Forcasting Peek Flood In The Madarsoo Basin Using Neural Network And Variable Several Regressions Method (Case Study: Madar Soo Basin). Geography, 11: 38. 113-132. (In Persian)
30. Poul, A.K., Shourian, M., and Ebrahimi, H. 2019. A comparative study of MLR, KNN, ANN and ANFIS models with wavelet transform in monthly stream flow prediction. Water Resources Management, 33: 8. 2907-2923.
31. Salahi, B., and Sarmasti, T. 2014. Simulation of Runoff-runoff presses in southern Sub-basin of Gharesou by Artificial neural networks mode (ANNS). Geography and environmental planning, 24: 4. 119-134. (In Persian)

32. Salarijazi, M., Ghorbani, K., Sohrabian, E., and Abdolhosseini, M. 2016. Prediction of Daily Stream-flow Using Data Driven Models. Iranian Journal of Irrigation & Drainage, 10: 4. 479-488. (In Persian)
33. Samadi, M., Bahremand, A., and Fathabadi, A. 2020. The Boustan Dam monthly inflow forecasting using data-driven and ensemble models in the Golestan Province, Watershed Engineering and Management, 11: 4. 1044-1058. (In Persian)
34. Solomatine, D.P., and Dulal, K.N. 2003. Model trees as an alternative to neural networks in rainfall-runoff modelling. Hydrological Sciences Journal, 48: 3. 399-411.
35. Sudheer, K.P., Gosain, A.K., and Ramasastri, K.S. 2002. A data-driven algorithm for constructing artificial neural network rainfall-runoff models. Hydrological processes, 16: 6. 1325-1330.
36. Talebi, A., and Akbari, Z. 2013. Investigation of ability of decision trees model to estimate river suspended sediment (case study: Ilam Dam basin). Journal of Science and Technology of Agriculture and Natural Resources, 17: 63. 109-121. (In Persian)
37. Valipour, M., Banihabib, M.E., and Behbahani, S.M.R. 2013. Comparison of the ARMA, ARIMA, and the autoregressive artificial neural network models in forecasting the monthly inflow of Dez dam reservoir. Journal of hydrology, 476: 433-441.
38. Vyas, S.K., Mathur, Y.P., Sharma, G., and Chandwani, V. 2016. December. Rainfall-Runoff Modelling: Conventional regression and Artificial Neural Networks approach. In 2016 International Conference on Recent Advances and Innovations in Engineering (ICRAIE) (pp. 1-6). IEEE.
39. Wang, W.C., Chau, K.W., Xu, D.M., and Chen, X.Y. 2015. Improving forecasting accuracy of annual runoff time series using ARIMA based on EEMD decomposition. Water Resources Management, 29: 8. 2655-2675.
40. Witten, I.H., and Frank, E. 2002. Data mining: practical machine learning tools and techniques with Java implementations. Acm Sigmod Record, 31: 1. 76-77.
41. Wood, E.F., and Rodríguez-Iturbe, I. 1975. Bayesian inference and decision making for extreme hydrologic events. Water Resources Research, 11: 4. 533-542.
42. Wu, C.L., and Chau, K.W. 2010. Data-driven models for monthly streamflow time series prediction. Engineering Applications of Artificial Intelligence, 23: 8. 1350-1367.
43. Zahiri, A.R., and Ghorbani, K.H. 2013. Flow discharge prediction in compound channels by using decision model tree M5. Journal of Water and Soil Conservation, 20: 3. 113-132. (In Persian)