Iranian Journal of Soil and Water Research ISNN: 2423-7833
Homepage: http://ijswr.ut.ac.ir

Improving the Estimation of Simulated River Discharge Values Using State Space Structural Models

Amin Mohammadzadeh shobegar[1], Mohammad Reza Sharifi [2]✉, Fereydoon Radmanesh [3], Behzad Mansouri[4]

1. Department of Hydrology and Water Resources, Faculty of Water and Environmental Engineering, Shahid Chamran University of Ahvaz, Ahvaz, Iran, Email: amin.mohammadzadeh63@gmail.com
2. Corresponding Author, Department of Hydrology and Water Resources, Faculty of Water and Environmental Engineering, Shahid Chamran University of Ahvaz, Ahvaz, Iran, Email: msharifi@scu.ac.ir
3. Department of Hydrology and Water Resources, Faculty of Water and Environmental Engineering, Shahid Chamran University of Ahvaz, Ahvaz, Iran, Email: radmanash@scu.ac.ir
4. Department of Statistics, Faculty of Mathematics and Computer Science, Shahid Chamran University of Ahvaz, Ahvaz, Iran, Email: b.mansoure@scu.ac.ir

Article Info

Article type: Research Article

Article history:

Received: June. 22, 2022

Revised: Sep. 9, 2022

Accepted: Sep. 13, 2022

Published online: Oct. 23, 2022

Keywords:
Time series,
State space models,
Ensemble method,
Monthly discharge,
Dez catchments

ABSTRACT

System simulation is done with different structures and by using different approaches and algorithms. Algorithms are intelligent methods of data processing in machine learning that can identify unknown factors in a time-dependent phenomenon. In the analysis of random phenomena, among the methods that can make decision-making easier is the ensemble algorithms. With the help of this method, more accurate data management and more knowledge of the studied system is obtained. Since, investigation of the trend component can be effective in simulating hydrological phenomena and help in interpreting the relationship between hydrological processes and environmental changes in the study areas; State space models have the advantage of analyzing the system flexibly and dynamically. Therefore, this article aims to improve the efficiency of Kalman Filter, ETS, BATS, and TBATS state space time series models with the help of an ensemble method and by comparing with the Box-Jenkins model, to show which of these models has a better capability in simulating the monthly discharge of the river. This comparison has been done in three water measuring stations of Sepiddasht Cesar, TangPanj Bakhtiari and Telezang in Dez catchments located in Khuzestan province since 1386 to 1399. The results of this study, based on the model evaluation criteria (RMSE, MAE and R^2), showed that the state space performed better than the Box-Jenkins model (classical), and among the state space models, the local level model (Kalman filter) performed better. So that in the validation stage, RMSE = 39.21and R2 = 0.79 in Sepiddasht Cesar water measuring station, RMSE = 57.89 and R2 = 0.76 in TangPanj Bakhtiari station and RMSE = 113.41 and R2= 0.73 in Telezang station were obtained.

Cite this article: Mohammadzadeh Shobegar, A., Sharifi, M. R., Radmanesh, F., & Mansouri, B. (2022). Improving the estimation of simulated river discharge values using state space structural models. *Iranian Journal of Soil and Water Research*, 53 (8), 1921-1936.

© The Author(s).

Publisher: University of Tehran Press.

DOI: http//doi.org/ 10.22059/ijswr.2022.344880.669299

مجله تحقیقات آب و خاک ایران، دوره ۵۳، شماره ۸ شاپا: ۷۸۳۳-۲۴۲۳
Homepage: http://ijswr.ut.ac.ir

بهبود برآورد مقادیر شبیه‌سازی‌شده دبی رودخانه با استفاده از مدل‌های ساختاری فضای حالت

امین محمدزاده شعبه‌گر[۱]، محمدرضا شریفی[۲]✉، فریدون رادمنش[۳]، بهزاد منصوری[۴]

۱. گروه هیدرولوژی و منابع آب، دانشکده مهندسی آب و محیطزیست، دانشگاه شهید چمران اهواز، اهواز، ایران، ایمیل: amin.mohammadzadeh63@gmail.com

۲. نویسنده مسئول، گروه هیدرولوژی و منابع آب، دانشکده مهندسی آب و محیطزیست، دانشگاه شهید چمران اهواز، اهواز، ایران، ایمیل: msharifi@scu.ac.ir

۳. گروه هیدرولوژی و منابع آب، دانشکده مهندسی آب و محیطزیست، دانشگاه شهید چمران اهواز، اهواز، ایران، ایمیل: radmanash@scu.ac.ir

۴. گروه آمار،دانشکده علوم ریاضی و کامپیوتر، دانشگاه شهید چمران اهواز، اهواز، ایران، ایمیل: b.mansoure@scu.ac.ir

اطلاعات مقاله

نوع مقاله: مقالهٔ پژوهشی

تاریخ دریافت: ۱۴۰۱/۴/۱
تاریخ بازنگری: ۱۴۰۱/۶/۱۸
تاریخ پذیرش: ۱۴۰۱/۶/۲۲
تاریخ انتشار: ۱۴۰۱/۸/۱

واژه‌های کلیدی:
سری زمانی،
مدل‌های فضای حالت،
روش ترکیبی،
دبی ماهانه،
حوضه آبریز دز.

چکیده

شبیه‌سازی سامانه، با ساختارهای متفاوت و با استفاده از رویکردها و الگوریتم‌های مختلف صورت می‌گیرد. الگوریتم ها روش‌های هوشمند پردازش داده در یادگیری ماشین هستند که می‌توانند عوامل ناشناخته در یک پدیده وابسته به زمان را شناسایی نمایند. در تحلیل پدیده‌های تصادفی از جمله روش‌هایی که می‌تواند تصمیم‌گیری را ساده‌تر کند، استفاده از الگوریتم‌های ترکیبی است. به‌کمک این روش، مدیریت داده دقیق‌تر و شناخت بیشتری از سامانه مورد مطالعه بدست می‌آید. از آنجایی‌که بررسی مؤلفه روند می‌تواند در شبیه‌سازی پدیده‌های هیدرولوژیکی مؤثر باشد و در تفسیر رابطه بین فرآیندهای هیدرولوژیکی و تغییرات محیطی در مناطق مورد مطالعه کمک مؤثری نماید؛ مدل‌های فضای حالت این مزیت را دارند که سامانه را به‌صورت انعطاف‌پذیر و پویا مورد بررسی و تحلیل قرار دهند. لذا این مقاله در نظر دارد به‌کمک روش ترکیبی به بهبود راندمان مدل‌های سری زمانی فضای حالت Kalman Filter، ETS، BATS،TBATS بپردازد و با مقایسه با مدل باکس–جنکینز نشان دهد کدامیک از این مدل‌ها، قابلیت بهتری در شبیه‌سازی دبی ماهانه رودخانه دارد. این مقایسه در سه ایستگاه آب‌سنجی سپیددشت سزار، تنگ‌پنج بختیاری و تله‌زنگ در حوضه آبریز دز واقع در استان خوزستان از سال ۱۳۸۶تا ۱۳۹۹ انجام شده است. نتایج این بررسی براساس معیارهای ارزیابی مدل(RMSE، MAE و R^2)، نشان داد فضای حالت نسبت به مدل باکس–جنکینز (کلاسیک) بهتر عمل نموده و در بین مدل‌های فضای حالت، مدل سطح موضعی(فیلتر کالمن) عملکرد بهتری داشته، به‌طوری‌که در مرحله صحت‌سنجی، ایستگاه آب‌سنجی سپیددشت سزار RMSE=۳۹/۲۱، R^2=۰/۷۹ و در ایستگاه تنگ‌پنج بختیاری RMSE= ۵۷/۸۹، R^2=۰/۷۶ و در ایستگاه تله‌زنگ RMSE=۱۱۳/۴۱ و R^2=۰/۷۳ بدست آمد.

استناد: محمدزاده شعبه‌گر، امین؛ شریفی، محمدرضا؛ رادمنش، فریدون؛ منصوری، بهزاد. (۱۴۰۱). بهبود برآورد مقادیر شبیه‌سازی‌شده دبی رودخانه با استفاده از مدل‌های ساختاری فضای حالت. *مجله تحقیقات آب و خاک ایران*، ۵۳ (۸)، ۱۹۳۶- ۱۹۲۱.

DOI: http//doi.org/10.22059/ijswr.2022.344880.669299

ناشر: مؤسسه انتشارات دانشگاه تهران.

© نویسندگان.

مقدمه

پیش‌بینی و برآورد جریان رودخانه یکی از مهم‌ترین جنبه‌های مطالعات هیدرولوژی است. نتایج این پیش‌بینی، مسئولان را قادر می‌سازد نسبت به کاهش خسارات ایجاد شده اقدامات احتیاطی را از قبل برای کنترل و پیشگیری سیل انجام دهند (Wu *et al.*, 2012). اکثر روش‌های پیش‌بینی نوعی شبیه‌سازی وضعیت سامانه است که از آن به‌عنوان مدل‌سازی نیز یاد می‌شود. این مدل‌ها را می‌توان به دو دسته عمده تقسیم کرد: مدل‌های مفهومی و مدل‌های مبتنی بر داده (Nayak *et al.*, 2004). از آنجایی‌که مدل‌های مفهومی به داده و دانش کافی از فرآیندهای مؤثر بر پدیده نیاز دارند، اغلب محققان به مدل‌های آماری مبتنی بر داده روی آورده‌اند. در دهه‌های اخیر، مدل سری زمانی به‌طور گسترده به‌عنوان یک مدل آماری در مطالعات اقلیمی، هیدرولوژی و منابع آب مورد استفاده قرار گرفته است (El-Shafie *et al.*,2007). فضای حالت یک رویکرد جدید در شناسایی فرآیندهای تصادفی است. این روش بر این فرض استوار است که سری زمانی به‌صورت ساختارهای جداگانه مدل‌بندی می‌گردد. باید توجه داشت که مؤلفه تصادفی در الگوبندی مدل الزامی است اما حضور یا عدم حضور دیگر مؤلفه‌ها با توجه به سری تحت مطالعه تعیین می‌شود. فضای حالت برای بروز کردن و برآورد پارامترهای مدل از الگوریتم‌های بهینه استفاده می‌کند. بهینه‌سازی مدل‌های هیدرولوژیکی با استفاده از فیلتر کالمن در دهه گذشته مورد توجه زیادی قرار گرفته است. کاربرد این الگوریتم زمانی ساده است که رابطه بین متغیرهای حالت و مشاهدات خطی باشد، به‌عبارت دیگر، زمانی که مشاهدات را بتوان مستقیماً در فضای حالت نگاشت کرد. به‌عنوان نمونه Ngan (1987)، مدل‌های خودرگرسیون را برای پیش‌بینی جریان مبتنی بر فضای حالت در رساله دکتری خود مقایسه کرد. وی نشان داد که فضای حالت در مقایسه با ARMAX از قابلیت اطمینان بهتری در پیش‌بینی جریان رودخانه برخوردار است. Sun *et al.* (2020) بهبود مهارت پیش‌بینی مدل‌های هیدرولوژیکی را با استفاده از مدل‌های ترکیبی فیلتر کالمن و فیلتر کالمن بی اثر[1] مورد بررسی قرار دادند و به این نتیجه رسیدند که در شبیه‌سازی دبی مشاهداتی مدل فیلتر کالمن بی‌اثر نسبت به مدل ترکیبی فیلتر کالمن (EnKF) عملکرد بهتری دارد. فیلترکالمن در حقیقت روشی انطباقی برای برآورد ضرایب در مدل‌های فضای حالت است که در بسیاری از مدل‌های خطی و غیر خطی مورد استفاده قرار می‌گیرد. Zeng *et al.* (2016) به تجزیه و تحلیل روندهای سری زمانی برای پیش‌بینی کوتاه مدت بیماری اپیدمی سیاه سرفه در کشور چین به کمک مدل‌های نمایی فضای حالت ETS (M, N, M) و مدل ARIMA پرداختند. نتایج آن‌ها نشان داد مدل ETS(M,N,M) با حداقل۵۲/۲۰۲= RSME در واسنجی و صحت سنجی بالاترین کارایی را دارد. De Livera *et al.* (2011) به‌بررسی پیش‌بینی سری‌های زمانی با الگوهای پیچیده فصلی و با استفاده از مدل‌های هموارسازی نمایی پرداختند؛ و نشان دادند که مدل BATS و TBATS واضح‌ترین تعمیم مدل‌های جدید فصلی است و می‌تواند بهتر از مدل SARIMA عمل کند. Lu *et al.* (2022) کاربرد روش تخمین پارامتر را در مدل هیدرولوژیکی بر اساس مدل ترکیبی فیلتر کالمن مورد بررسی قرار دادند و نتایج آن‌ها نشان داد که کمبود داده آب و هواشناسی اگر چه بر دقت پارامتر در شبیه‌سازی جریان تأثیر می‌گذارد اما مدل ترکیبی فیلتر کالمن(EnKF) می‌تواند به‌عنوان یک روش تصحیح پارامتر برای حل این مشکل استفاده شود. Shen *et al.* (2022)توسعه سامانه مدل‌سازی هیدرولوژیکی را به‌کمک مدل ترکیبی فیلتر کالمن جهت پیش‌بینی جریان رودخانه مورد بررسی قرار دادند و نشان دادند که مدل هیدرولوژیکی (SHMS) بر روی مدل ترکیبی فیلتر کالمن (EnKF) باعث بهبود تخمین پارامترهای مدل هیدرولوژیکی از طریق همگرایی سریع‌تر و دقیق‌تر در شبیه‌سازی جریان رودخانه شده است. Stefenon *et al.* (2022) پیش‌بینی سری‌های زمانی با استفاده از روش‌های یادگیری ترکیبی برای مواقع اضطراری در نیروگاه برق آبی مورد بررسی قرار دادند؛ آن‌ها برای داشتن یک سامانه تصمیم‌گیری سریع‌تر از مدل‌های مختلف یادگیری ترکیبی استفاده کردند؛ نتایج آن‌ها نشان داد مدل‌های ترکیبی باعث افزایش دقت مدل شده است. یادگیری ترکیبی که یک الگوی یادگیری ماشینی است و در آن چندین مدل که اغلب یادگیرنده ضعیف[2] نامیده می‌شوند برای حل یک مشکل، آموزش داده شده و برای به‌دست آوردن نتایج بهتر ترکیب می‌شوند. فرضیه اصلی این است که وقتی مدل‌های ضعیف به‌درستی ترکیب شوند، مدل‌های دقیق‌تر و یا قوی‌تری به‌دست می‌آید (Dong *et al.*, 2020). از دیگر نکات مهم در آموزش مدل، گزینش و بهینه‌سازی نرخ آموزش و در نهایت شروط توقف آموزش است. بهینه‌سازی به دو دسته محلی و سراسری تقسیم‌بندی می‌شود. بدیهی است اختلاف زیادی بین کمینه سراسری و کمینه‌های محلی وجود دارد. اما اگر به‌طور تصادفی در یک الگوریتم بهینه‌سازی، منطقه‌ای را شناسایی نماییم، شاید بر کمینه کلی نیز منطبق شود. بنابراین مناسب است که در عمل و جهت بهینه‌سازی هر سامانه با ساختار یکسان را، چندین مرتبه آموزش داد تا در نهایت عدم‌قطعیت حاصل از الگوریتم‌های بهینه‌ساز محلی، مورد بررسی قرار

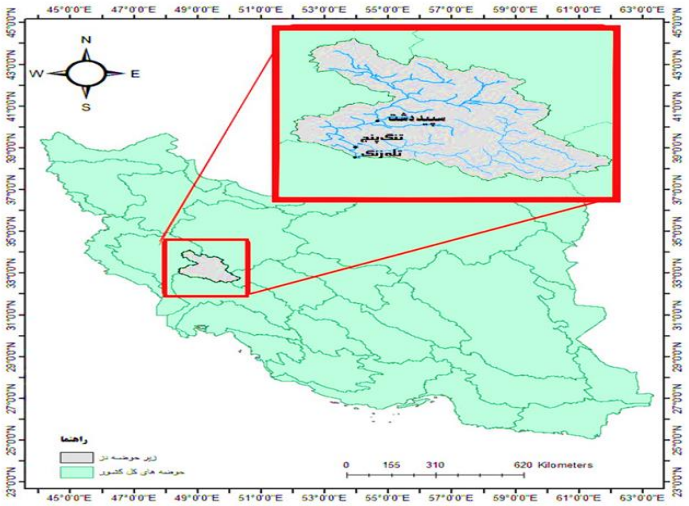
1 UKF: Unscented Kalman Filter
2 Weak learner

گیرد (Fletcher, 1980). به‌طور کلی در یادگیری ماشین، انتخاب مدل برای داشتن هر گونه شانسی برای به‌دست آوردن نتایج خوب بسیار مهم است. این انتخاب می‌تواند به بسیاری از متغیرهای مسئله بستگی داشته باشد مانند: کمیت داده‌ها، ابعاد فضا، فرضیه توزیع و غیره. واریانس کم، یکی از ویژگی اساسی مورد انتظار برای یک مدل است که این کار به‌کمک مدل‌های ترکیبی انجام می‌گیرد. برای راه‌اندازی یک روش یادگیری ترکیبی، از یک الگوریتم یادگیری پایه استفاده می‌شود تا با یادگیرندگان ضعیف همگنی داشته باشند (Pintelas & Livieris, 2020). الگوریتم‌های ترکیبی انواع مختلفی دارند که هدف آن‌ها ترکیب یادگیرندگان ضعیف و کاهش عدم‌قطعیت مدل‌ها است که در این پژوهش از روش بگینگ استفاده شده است (Liu, 2009). لذا با توجه به مطالعات انجام شده این مقاله در نظر دارد به بررسی ساختاری مدل سری زمانی با استفاده از رویکرد جدید آماری به نام فضای حالت، که با ساختارهای متفاوت (Kalman Filter، ETS، BATS، TBATS) و با فرضیات کمتر، به شناسایی واقعیت سامانه بپردازد و به این پرسش پاسخ دهد که ورود متغیر تصادفی دبی در ساختارهای متفاوت فضای حالت چگونه موجب بهبود نتایج شبیه‌سازی می‌شود؟ به بیان دیگر با توجه به در اختیار بودن داده‌های ماهانه آیا روش ترکیبی، کمکی به ارتقاء مدل سری زمانی فضای حالت نسبت به کلاسیک (باکس–جنکینز) می‌کند یا خیر؟ لازم به ذکر است این بررسی در سه ایستگاه آب‌سنجی (سپید دشت سزار، تنگ‌پنج بختیاری و تله‌زنگ) در حوضه آبریز دز واقع در استان خوزستان صورت پذیرفته است.

مواد و روش‌ها

منطقه مورد مطالعه

حوضهٔ آبریز رودخانه دز به‌عنوان یک حوضهٔ درجهٔ سه، زیر مجموعه‌ای از حوضهٔ کارون بزرگ محسوب می‌شود و در تقسیم‌بندی بزرگتر در زیر مجموعهٔ حوضهٔ خلیج فارس و دریای عمان قرار می‌گیرد. مساحت کل حوضه بالغ بر ۲۱۷۲۰ کیلومترمربع و متوسط ارتفاع حوضه حدود ۱۶۰۰متر می‌باشد. محدودهٔ مورد مطالعه در این پژوهش حوضهٔ آبریز دز است که در بالادست سد دز و مساحت آن حدود ۱۷۳۶۵ کیلومترمربع است. شیب حوضه در بالادست سد نسبتاً تند بوده به‌طوری‌که شیب متوسط آن ۱۲٫۱ درصد می‌باشد و ۱۰ درصد حوضه شیبی بیشتر از ۱۹٫۵ درصد دارد. شیب عمومی و جهت جریان در حوضهٔ آبریز دز از سمت شمال به جنوب است. رودخانهٔ دز در ابتدا از اتصال دو شاخهٔ اصلی رودخانه تیره با رودخانهٔ ماربره در محل شهر درود که رودخانهٔ سزار را تشکیل می‌دهد؛ شروع شده و با دریافت سر شاخه‌های سبزه، واسک، زاز و سرخاب، نهایتاً رودخانهٔ دز در محلی به‌نام تنگ‌پنج در اعماق رشته کوه زاگرس از تلاقی دو رودخانهٔ سزار و بختیاری تشکیل می‌شود. در حوضهٔ آبریز دز، ۳۶ ایستگاه آب‌سنجی وجود دارد که از این تعداد، ۲۹ ایستگاه در بالادست سد دز و ۷ ایستگاه در پایین‌دست سد دز واقع شده است. توزیع ایستگاه‌های آب‌سنجی در شبکهٔ رودخانهٔ دز، به‌گونه‌ای طراحی شده است که قبل از تلاقی هر شاخهٔ مهم با شاخهٔ اصلی رودخانهٔ دز، یک ایستگاه آب‌سنجی بر روی سرشاخه احداث گردیده است. در این مقاله، به‌منظور تحلیل سری‌های زمانی حوضهٔ آبریز دز از داده‌های ماهانه دبی شرکت مدیریت منابع آب کشور و سازمان آب و برق خوزستان از سال (۱۳۸۶–۱۳۹۹) استفاده شده است. در جدول(۱) موقعیت ایستگاه‌ها و در شکل(۱) محدوده جغرافیایی ایستگاه‌های منتخب نشان شده است.



شکل ۱- محدودهٔ مطالعاتی حوضهٔ آبریز دز و ایستگاه‌های منتخب آب‌سنجی

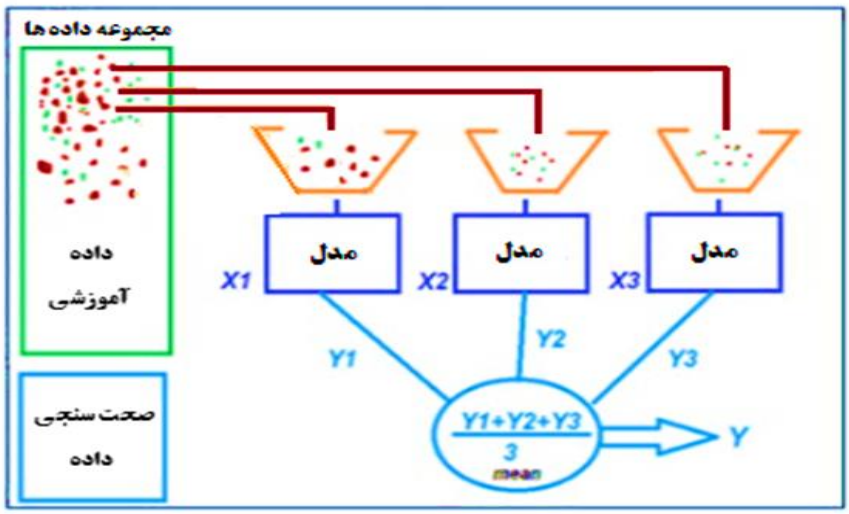
جدول ۱- مشخصات ایستگاه‌های آب‌سنجی منتخب در حوضهٔ آبریز دز.

ردیف	رودخانه	ایستگاه	طول جغرافیایی	عرض جغرافیایی
۱	سزار	سپید دشت	۴۸٫۸۹	۳۳٫۲۲
۲	بختیاری	تنگ پنج	۴۸٫۷۶	۳۲٫۹۳
۳	دز	تله زنگ	۴۸٫۷۷	۳۲٫۸۲

روش‌شناسی در سری‌های زمانی

معرفی تئوری بگینگ[1]

این الگوریتم توسط لئوبریمن در سال ۱۹۹۶ مطرح شد (Breiman, 1996). بگینگ به‌منظور بهبود یادگیری ضعیف ماشین مورد استفاده قرار می‌گیرد. بگینگ مخفف «جمع‌بندی راه‌انداز» است که هدف آن تولید یک مدل ترکیبی است که قوی‌تر (مقاوم‌تر) از مدل‌های مجزای سازنده آن عمل کند. ایده بگینگ ساده است و می‌تواند چندین مدل مستقل را برازش نموده و پیش‌بینی‌های آن‌ها را میانگین گرفته تا مدلی با واریانس کمتر به‌دست آید. هنگام آموزش یک مدل، مشکل طبقه‌بندی یا رگرسیون مهم نیست؛ زیرا تابعی به‌دست می‌آید که یک ورودی می‌گیرد و یک خروجی بر می‌گرداند و با توجه به کل داده‌ها، داده‌های آموزشی تعریف شده و نیز با توجه به واریانس داده آموزشی، مدل برازش شده مشمول تنوع می‌گردد؛ و اگر داده جدیدی مشاهده شد، یک مدل متفاوت دیگری به‌دست می‌آید. از مزیت‌های این الگوریتم آن است که برای دستیابی به نتیجه بهینه تنها نیاز به تنظیم یک پارامتر دارد و به‌طور موازی عمل می‌کند. در روش‌های موازی، یادگیری‌های مختلف در نظر گرفته شده را، مستقل از یکدیگر برازش نموده و بنابراین امکان آموزش همزمان آن‌ها وجود دارد. در این روش عملاً از ترفند بوت استرپ[2] استفاده می‌شود. واژه بوت استرپ به‌معنای نمونه‌برداری تصادفی با جایگزینی از نمونه‌های اصلی با همان اندازه است (Huang *et al.*,2009). در این روش داده می‌تواند از مجموعه داده‌ها چند بار انتخاب یا اصلاً انتخاب نشود، ولی سایز آموزش مدل، برابر همان سایزی است که در نظر گرفته می‌شود (Dantas *et al.*, 2018). به‌طور کلی داده‌ها به دو دسته آموزش و تست تقسیم‌بندی می‌شوند. همان‌طور که در شکل (۲) مشاهده می‌شود به‌کمک داده‌های آموزش سه بوت استرپ ایجاد می‌گردد که با پارامتر t نمایش داده شده است(t=3)؛ که بدین معنی است سه بار نمونه‌برداری باجایگزینی انجام می‌گیرد. اگر این روش در بخش رگرسیون استفاده شد؛ داده جدیدی که در مرحله تست به‌کار گرفته می‌شود با توجه به مدل رگرسیون خروجی، تحت عنوان y_1 ،y_2 و y_3 تخمین زده می‌شود و میانگین آن‌ها، y نهایی را برای ما تولید می‌کند. برای توضیحات بیشتر می‌توان به مطالعات Bühlmann *etal.* (2002) و Bühlmann (2012) مراجعه نمود.



شکل ۲- نمایش الگوریتم بگینگ

1 Bagging
2 Bootstrap aggregating

معرفی رویکرد باکس-جنکینز (کلاسیک)

یکی از روش‌های رایج در مدل‌سازی متغیرهای هیدرولوژی رویکرد باکس-جنکینز است؛ که به‌صورت مفهومی توضیح داده شده و از عبارات ریاضی رایج در این مدل خودداری گردیده است. مدل آریما در واقع یک سر واژه است که خلاصه مدل میانگین متحرک خود بازگشتی یکپارچه[1] می‌باشد. این مدل از سه پارامتر اصلی p و d و q تشکیل شده است. پارامتر p درجهٔ خود بازگشتی مدل و پارامتر d یکپارچگی مدل و پارامتر q مربوط به میانگین متحرک می‌باشد. منظور از مؤلفهٔ خود بازگشتی، در مدل‌های رگرسیونی در واقع مقادیر قبلی، مقدار کنونی است. خودبازگشتی به معنای یک رگرسیون از خودش است که می‌تواند هر متغیری باشد و متغیر دیگر در حقیقت یک متغیر وابسته به زمان است که با y(t) نشان داده می‌شود. این مقدار y(t) بستگی به مقدار قبلی خود دارد. نکته اصلی آن است که این متغیر یک مؤلفه رگرسیونی است که روی خودش رگرس می‌شود و به مقادیر پیشین وابسته است. پارامتر d درجهٔ تفاضل‌گیری را نشان داده که منظور از تفاضل‌گیری، تفاضل‌گیری غیرفصلی است. پارامتر q میانگین متحرک که در واقع خطای مدل را نشان می‌دهد. به‌همین علت بعضی وقت‌ها با e(t) هم نمایش داده می‌شود. پیش فرض این مدل آن است که داده‌ها مانا شوند. یعنی روندها و فصلی بودن در داده‌ها اصلاح گردند. برای آزمون مانایی داده‌ها، دو مؤلفهٔ مهم را باید در نظرگرفت: ۱-تابع خود همبستگی (ACF)،۲-تابع خود همبستگی جزئی (PACF). خود همبستگی به‌معنای شباهت میان مقادیر یک متغیر در مشاهدات مختلف است. با رسم نمودار تابع خود همبستگی، همبستگی داده‌ها نشان داده می‌شود؛ مقدار تابع خود همبستگی یک عددی بین ۱- و ۱ است. اما تابع خود همبستگی جزئی، درجهٔ رابطه بین دو متغیر را بیان می‌کند. در سال‌های گذشته تمرکز بسیاری از تحقیقات در زمینه سری‌های زمانی منوط به ارائه روش‌های نوین در برآورد مدل و بهبود نتایج پیش‌بینی صورت گرفته است؛ در حالی که بهبود فرآیند شناسایی ساختار مدل نیز می‌تواند بر سرعت دستیابی به مدل بهینه مؤثر باشد، زیرا انتخاب مدل نادرست می‌تواند باعث افزایش هزینهٔ زمان دستیابی به مدل بهینه شود (Box & Jenkins, 1970). در نرم افزار R بسته‌ای وجود دارد که می‌تواند بدون آن‌که با مفهوم کلروگرام آشنا باشیم پارامترهای p و d و q مدل را به‌صورت اتوماتیک بدست آورد. با دستور auto. arima این کار به‌سادگی امکان‌پذیر بوده و می‌تواند تحلیلی کاوشگرانه ارائه دهد و با یک آزمون آماری به نام لیانگ – باکس[2] مدل را اعتبارسنجی[3] کند. برای توضیحات بیشتر به مطالعات Makridakis *et al.* (1997) می‌توان مراجعه نمود.

معرفی رویکرد مدل‌سازی فضای حالت

مدل خود رگرسیون فضای حالت از دو معادلهٔ حالت و اندازه‌گیری تشکیل شده است (Shumway & Stoffer,1982; Shumway, 1988). طریقهٔ ارتباط این دو معادله بدین گونه است:

$$Z_i = \Phi Z_{i-1} + \omega_i \quad \text{رابطه ۱)}$$

$$X_i = M_i Z_i + \nu_i \quad \text{رابطه ۲)}$$

برای سری زمانی X_i، معادله حالت مرتبه اول که معمولاً استفاده می‌شود، بردار حالت Z_i را در زمان i نسبت به حالت Z_{i-1} در زمان قبلی i-1 را توصیف می‌کند: که در آن Φ ماتریس انتقال و ω_i خطای مدل است. با در نظر گرفتن عدم‌قطعیت اندازه‌گیری علاوه بر خطای مدل، مدل اندازه‌گیری بردار X_i مشاهده شده را از طریق ماتریس M_i به بردار حالت واقعی Z_i و خطای اندازه گیری ν_i مرتبط می‌کند. از آنجایی‌که دو ماتریس Φ و M_i و توزیع دو ماتریس خطا ناشناخته هستند، روش رایج حداقل مربعات برای تخمین مدل قابل استفاده نیست. در عوض، استنتاج مبتنی بر احتمال به‌نام فیلتر کالمن (Kalman,1960) اجازه ساخت تابع درستنمایی مرتبط با یک مدل فضای حالت را می‌دهد. از دیدگاه ریاضی، رویکرد فضای حالت، استفاده از روش‌های ماتریسی و برداری برای مدیریت تعداد زیادی از متغیرهایی است که وارد چنین مسئله‌هایی می‌شوند. به این ترتیب، (Chan & Palma ,1998) بیان نمودند، این‌ها روش‌های جدیدی نیستند، بلکه آن‌ها کشف مجدد تکنیک‌های ریاضی موجود هستند. از دیدگاه فیزیکی، رویکرد فضای حالت به تفکر مفهومی کمک می‌کند و مبنایی واحد برای حل مسائل خطی و غیرخطی می‌باشد. هم‌چنین این رویکرد برای تجزیه و تحلیل سامانه‌های یکپارچه و پویا، مناسب است. در ادامه مختصری از برخی از زیر کلاس‌های فضای حالت شامل مدل سطح موضعی با اثرات فصلی (فیلتر کالمن)، ETS، BATS و TBATS که در شبیه‌سازی دبی ماهانه به‌کار گرفته شده، توضیح داده می‌شود.

1Auto Regressive Integrated Moving Average
2 Ljung- box
3 Validation

معرفی مدل‌های سری زمانی فضای حالت

مدل‌های سری زمانی دارای چهار جزء اصلی است: فصلی، روند، دوره‌ای و تصادفی. با توجه به این‌که داده‌های آب و هوایی مطمئناً از تغییرات فصلی پیروی می‌کند و اوج و فرود منظم دارد؛ و تقریباً نوسانات فصلی از نظر اندازه در طول زمان ثابت است به‌همین دلیل از مدل‌های جمعی فضای حالت استفاده شده که در زیر به آن‌ها اشاره می‌گردد. در ضمن، جزئیات روابط مدل‌های فضای حالت در مطالعات Sharifi *et al.* (2022) بیان شده است.

-مدل سطح موضعی (فیلتر کالمن)

این مدل مناسب برای کاربرد در یک سری زمانی غیر ایستا است که می‌توان از آن در سری‌های زمانی فصلی استفاده کرد. هدف اصلی به‌دست‌آوردن تخمین‌های بهینه از حالت در نقطه زمانی t، تنها با در نظر گرفتن مشاهدات $\{y_1, y_2, \ldots, y_{t-1}\}$ است. واژه بهینه به‌معنای به حداقل رساندن میانگین مربعات خطای تخمین تفسیر می‌گردد (Durbin & Koopman, 2012). منظور از فیلتر کردن این است که هر بار یک مشاهده جدید وارد سامانه می‌شود، می‌تواند دانش خود را بروز کند. فیلتر کالمن (Kalman,1960) اساساً الگوریتمی برای بازبینی ممان‌های مؤلفه‌های تصادفی یک مدل سری زمانی خطی است. اغلب به‌عنوان یک پله برای استخراج لحظه‌ای مؤلفه‌ها در مقاطع زمانی آینده برای اهداف پیش‌بینی استفاده می‌شود. بازبینی لحظه‌ها به‌صورت بازگشتی انجام می‌گیرد. هم‌چنین می‌توان از آن به‌عنوان فرآیند مورد نیاز برای یافتن برآوردهای بیشترین احتمال پارامترهای مدل استفاده کرد (Harvey, 1990).

-ETS

Peggles (1969) به‌دلیل تعریف طبقه‌بندی ETS که بعدها توسط Gardner (1985) و Hyndman etal., (2002) توسعه داده شد، شناخته می‌شود (Shamshad *etal.*, 2019). مدل‌های دینامیکی غیرخطی را با استفاده از مدل‌های ETS، که ترکیبی از مدل‌های هموارسازی نمایی است، می‌توان ساخت. این مدل یکی دیگر از مدل‌های سری زمانی فضای حالت می‌باشد که به کمک آن هموارسازی صورت گرفته و داده‌های مربوط به گذشته، رفته رفته نقش کمتری در محاسبهٔ میانگین خواهند داشت. این سیر نزولی به‌صورت نمایی تغییر کرده و در نتیجه به آن هموارسازی نمایی می‌گویند. این مدل پیش‌بینی کاملاً خودکار است و بدون دخالت انسان ویژگی‌های داده را استخراج می‌کند. زمانی‌که سطح، روند، فصل و بی‌نظمی(خطا) وجود دارد، برای پیش‌بینی داده‌های سری زمانی از مدل ETS استفاده می‌شود. این مدل انعطاف‌پذیر بوده و از خاصیت استحکام پیروی می‌کند و به‌عنوان مدلی برای برون‌یابی تلقی می‌شود(Zeng *et al.*,2016). کفایت مدل به کمک بررسی خودهمبستگی باقیمانده‌ها روشی معمول بوده که در مدل‌های فضای حالت معمولاً از آزمون لیانگ–باکس در مدل‌های خطی استفاده می‌شود. اطلاعات در خصوص ETS در پژوهش‌های Abd ;Alakkari *etal.*,2022; Jain and Mallick , 2017 Rahman *etal.*, 2017 آمده است.

-BATS & TBATS

BATS و TBATS دو نوآوری در چارچوب مدل‌سازی فضای حالت هستند. این مدل‌ها برای سری‌های زمانی فصلی پیچیده که شامل الگوهای فصلی متعدد، غیرصحیح و اثرات تقویم دوگانه است؛ کاربرد دارند. تغییرات فصلی برای تصمیم‌گیری در آینده مهم هستند. فصلی غیر صحیح، بدان معنی است که داده‌های فصلی ماهانه را نمی‌توان با ۳۰ مقدار صحیح نشان داد زیرا ماه‌های اسفند ۲۹ روز است در حالی‌که ماه‌های دیگر ۳۰ روز یا ۳۱ روز هستند. این منجر به طول متوسط ۳۰/۴ روز می‌شود. اما دوره‌های فصلی چندگانه به این معنی است که داده‌ها بیش از یک فصل دارند، مانند داده‌های سری زمانی روزانه که ممکن است فصلی، هفتگی و سالانه باشد. مفهوم تقویم دوگانه به نقش تقویم‌های مختلف و تأثیر آن‌ها بر پیش‌بینی روش‌های سری زمانی اشاره دارد مانند کشور هند که از تقویم هندی و میلادی پیروی می‌کند. برای مدل کردن چنین مشخصاتی از مدل‌های BATS و TBATS استفاده می‌شود (De Livera *et al.*,2011). مزایای ساختار این مدل عبارتند از: (۱) تخمین پارامتر بهبود یافته (۲) توانایی مدیریت مؤلفه پیچیده و ویژگی‌های ذاتی سری‌های زمانی. TBATS به‌عنوان یک مدل فضای حالت، یک فضای پارامتری متعدد را با هدف پیش‌بینی بهتر فراهم می‌کند (Hyndman *etal.*, 2008). اطلاعات در خصوص BATS و TBATS در پژوهش‌های Gos *etal.*, 2020; De Livera, 2010; Badr *etal.*,2021; Naim *etal.*,2018; Zachary *etal.*,2020 آمده است.

معیارهای ارزیابی مدل

استفاده درست از سری‌های زمانی بیشتر یک هنر است تا یک علم، بهترین راه برای فراگیری این هنر، تجربه و تکرار است. با توجه به این

که بیشتر اوقات در سری‌های زمانی برای شناسایی الگوی مناسب از روش‌های شهودی استفاده می‌شود، نباید تنها به یک روش برای شناسایی و انتخاب مدل اکتفا کرد، انتخاب مدل بر اساس توالی مراحل اجرای یک مدل سری زمانی و تکرار چند باره این توالی، سبب شناسایی مدل مناسب خواهد شد. یک ماشین، الگوها و خصوصیات مختلفی را از داده‌هایی که به آن آموزش داده شده است یاد می‌گیرد و خود را برای تصمیم‌گیری‌های مختلف مانند شناسایی سامانه یا پیش‌بینی داده‌های جدید، آموزش می‌دهد. برای بررسی این‌که چگونه ماشین قادر به اتخاذ این تصمیمات است؛ لازم است ابتدا بر روی داده‌های آموزش داده شده کار شود و پس از آموزش مدل به اندازه کافی، از آن برای تست مدل بر اساس داده‌های جدید استفاده گردد. صحت‌سنجی، تعیین میزان تطابق مدل با واقعیت است. یکی از راه‌های تشخیص صحت مدل، هم‌خوانی نتایج گرفته شده از مدل با واقعیت است. باید توجه کرد که تطابق صد در صد مدل با واقعیت غیر ممکن است. در ایده‌آل‌ترین حالت، فقط می‌توان ادعا کرد که مدل با واقعیت در تقابل نیست. لذا، به‌منظور ارزیابی مدل‌های مورداستفاده در این پژوهش از معیارهای ارزیابی میانگین خطای مطلق، ریشه میانگین مربعات خطا و ضریب تعیین که نشان‌دهندهٔ نیکویی برازندگی روند است به شرح زیر استفاده گردید.

رابطه ۳)
$$MAE = \frac{1}{n}\sum_{i=1}^{n}|Q_s\text{-}Q_o|$$

رابطه ۴)
$$RMSE= \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(Q_o\text{-}Q_s\right)^2}$$

رابطه ۵)
$$R^2 = \frac{[\sum_{i=1}^{n}(Q_o\text{-}\overline{Q_o})(Q_s\text{-}\overline{Q_s})]^2}{\sum_{i=1}^{n}(Q_o\text{-}\overline{Q_s})^2 \sum_{i=1}^{n}(Q_s\text{-}\overline{Q_s})}$$

که در آن‌ها، Q_o مقدار دبی مشاهده‌ای، Q_s دبی شبیه‌سازی‌شده، n تعداد داده‌ها، $\overline{Q_o}$ میانگین دبی‌های مشاهده‌ای، $\overline{Q_s}$ میانگین دبی‌های محاسباتی است(Huang *et al.*, 2022).

نتایج و بحث

مراحل یادگیری ماشین

در فرآیند داده‌کاوی و استخراج دانش، این داده‌ها هستند که نقش بسیار مهم و حیاتی دارند زیرا الگوریتم‌های مختلف بایستی بر روی داده‌ها، عملیات خود را انجام دهند. در واقع هر چه داده‌ها بهتر و مناسب‌تر باشد مدل یادگیری بهتر انجام می‌گیرد. پس بهتر است برای یادگیری مدل وارد حوزهٔ طبقه‌بندی داده شد. مسائل حوزه طبقه‌بندی که پایه آن آمار و ریاضیات است؛ یک خروجی مورد انتظار دارد و آن این است که بتوان برچسب‌ها را با دقت بالا شناسایی و از یکدیگر تفکیک کرد. در واقع برای این کار با مفهومی به‌نام بایاس و واریانس مواجه شده که این مفهوم به ما می‌گوید باید یک نقطه تعادل معقولی بین آن‌ها برقرار کرد (Yeh *etal.*, 1981). برای رسیدن به این نقطهٔ بهینه تعادل راهکارهای مختلفی وجود دارد یکی از راهکارها این است که داده‌های یادگیری ماشین به دو بخش آموزش و تست تقسیم گردند. طبقه‌بند بر اساس داده‌های آموزش ساخته می‌شود و بعد از آن یک ارزیابی با توجه به داده‌های جدا شده بخش تست انجام می‌گیرد تا از این طریق متوجه شد که آیا بایاس یا واریانس رخ داده است یا خیر؟ بایاس بالا نتیجهٔ سادگی طبقه‌بند است و واریانس بالا نتیجهٔ پیچیدگی بیش از حد طبقه‌بند است. طبقه‌بندهای ترکیبی از ترکیب چندین طبقه‌بند می‌باشند و طبقه‌بندها هر کدام مدل خود را بر روی داده‌ها ساخته و این مدل را ذخیره می‌کند. یکی از این طبقه‌بندها الگوریتم بگینگ است؛ این الگوریتم بیشتر برای ایجاد مدلی ترکیبی با واریانس کمتر تمرکز دارد و هدف آن ایجاد مدلی است که پایدارتر از مدل‌های پایهٔ سازندهٔ خود باشد. این الگوریتم به الگوریتم‌های غیر ثابت[1] نیز شناخته می‌شود. برای ایجاد یک روش یادگیری ماشین ترکیبی، ابتدا باید مدل‌های پایه را انتخاب و به صورت منطقی و منسجم با یکدیگر ترکیب کرد. به این معنی که اگر یک مدل پایه، با بایاس کم و واریانس بالا انتخاب شود از یک روش ترکیبی که تمایل به کاهش واریانس دارد استفاده می‌گردد (Kotsiantis *etal.*,2006). در این راستا افراد ذیل به نتایج مشابهی دست یافتند. Ribeiro & dos Santos Coelho (2020) روش‌های ترکیبی مبتنی بر بگینگ، بوستینگ، استکینگ را برای پیش‌بینی کوتاه مدت در سری‌های زمانی اقتصاد کشاورزی مورد بررسی قرار دادند؛ نتایج آن‌ها نشان داد که روش‌های ترکیبی دستاوردهای آماری قابل توجهی را نشان داده و

1Unstable

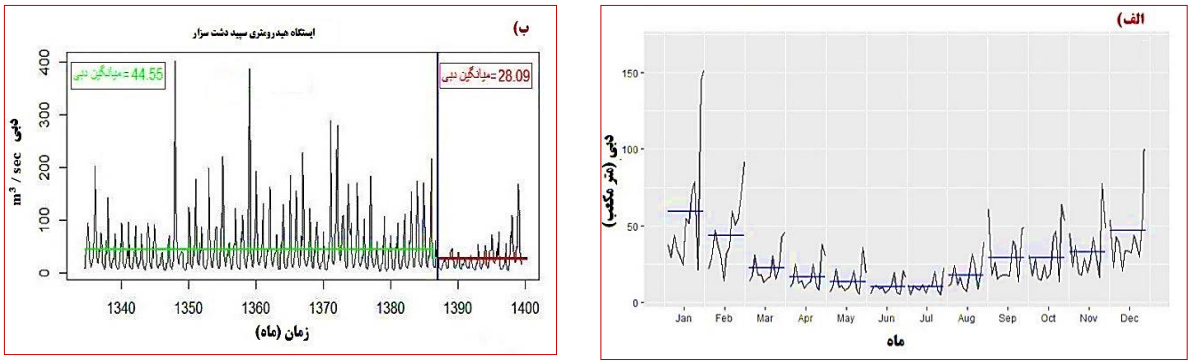
خطاهای پیش‌بینی را برای سری‌های زمانی کاهش می‌دهد لذا استفاده از روش‌های ترکیبی توصیه می‌شود. Gastinger *et al.* (2021) به پیش‌بینی سری‌های زمانی با استفاده از متا الگوریتم‌ها در بحث یادگیری ترکیبی پرداختند؛ نتایج آن‌ها نشان داد که روش‌های ترکیبی باعث دقت پیش‌بینی و بهبود مدل می‌گردد. Murray (2018) اهمیت روش‌های ترکیبی را برای پیش‌بینی آب و هوا مورد بررسی قرار داد؛ نتایج نشان داد که روش‌های ترکیبی باعث بهبود پیش‌بینی‌شده و می‌تواند در کاهش عدم‌قطعیت مدل مؤثر باشد. Seising & Allende-Cid (2017) مطالعات فازی و محاسبات نرم را مورد بررسی قرار دادند؛ نتایج آن‌ها نشان داد که روش‌های ترکیبی، نقاط قوت و ضعف مدل‌ها را در مجموع کاهش می‌دهد؛ و این روش دقت پیش‌بینی را تا حد زیادی افزایش می‌دهد. Wideroos (2021) بهبود دقت پیش‌بینی سری‌های زمانی را به‌کمک روش‌های ترکیبی مورد بررسی قرار داد؛ نتایج نشان داد که الگوریتم‌های ترکیبی می‌توانند دقت پیش‌بینی مدل‌های یادگیری ماشین را بهبود بخشند. Li *et al.* (2020) در مقاله‌ای با عنوان استراتژی ترکیبی به پیش‌بینی جریان ماهانه رودخانه یانگ‌تسه پرداختند؛ نتایج آن‌ها نشان داد که مدل‌های مبتنی بر یادگیری ماشین پتانسیل پیش‌بینی جریان ماهانه را بهبود می‌بخشد. Rajesh & Rehana (2021) به پیش‌بینی دمای آب رودخانه با استفاده از الگوریتم‌های یادگیری ماشین پرداختند؛ نتایج آن‌ها نشان داد که ضرایب عملکرد یادگیری ماشین در داده‌های ماهانه در مقایسه با مقیاس زمانی روزانه بهبود می‌یابد و مقیاس زمانی ماهانه در مدل‌های یادگیری ماشین نسبت به داده‌های روزانه و فصلی در این مطالعه بهتر عمل کرده است. Wang *et al.* (2015) بهبود دقت پیش‌بینی سری‌های زمانی رواناب سالانه با استفاده از مدل آریما بر اساس روش ترکیبی تجزیه، مورد بررسی قرار دادند؛ نتایج آن‌ها نشان داد که این مدل ترکیبی می‌تواند به‌طور قابل توجهی رویکردهای سری زمانی آریما را برای پیش‌بینی سری‌های زمانی سالانه بهبود بخشد.

Ardabili *et al.* (2019) به بررسی هدفمند مدل‌های یادگیری عمیق و یادگیری ماشین در تحقیقات سوخت‌های زیستی پرداختند؛ نتایج آن‌ها نشان داد که مدل‌های یادگیری اعم از مدل‌های هیبریدی و ترکیبی باعث افزایش دقت نسبت به مدل‌های مجزا می‌شود. در این پژوهش، برای شبیه‌سازی دقیق‌تر مدل‌های فضای حالت که تمایل بالایی در کاهش بایاس دارند؛ نیاز است برای کاهش واریانس از مدل‌های ترکیبی استفاده گردد تا آموزش مدل به‌خوبی انجام گیرد. رویکرد فضای حالت در دو نوع ساختار شروع به مدل‌سازی می‌کند: فیلتر کردن و هموار کردن؛ که هر کدام از آن‌ها پیش‌بینی خاص خود را ارائه می‌دهند. بررسی هر یک از این ساختارهای متفاوت، درک بهتری در شناسایی سامانه نشان می‌دهد. با این حال ساختار مدل به‌تنهایی نمی‌تواند کافی باشد چرا که ساختارهای متفاوتی در بحث سری زمانی وجود دارد. نکته این است که فضای حالت این امکان را می‌دهد که می‌توان در ساختارهای متفاوت و با فرضیات کمتر (ایستا کردن) شروع به مدل‌سازی کرد. در مدل‌سازی سری زمانی، دو تغییرات دارای اهمیت می‌باشند:۱- تغییرات در میانگین ۲- تغییرات در واریانس. فضای حالت به‌دلیل آن‌که مدلی پویا است و از الگوریتم‌های بهینه برای بدست آوردن پارامترها استفاده می‌کند؛ می‌تواند تغییرات در میانگین را به‌خوبی مدل‌سازی نماید ولی در کاهش واریانس دچار ضعف‌هایی است؛ به‌همین منظور جهت بهبود شبیه‌سازی دبی رودخانه از الگوریتم ترکیبی استفاده گردید. به‌طور کلی برای تحلیل نظام‌مند و خودکار برای ساخت مدل، مراحل یادگیری ماشین باید طی شود که در شکل (۳) نشان داده شده است.

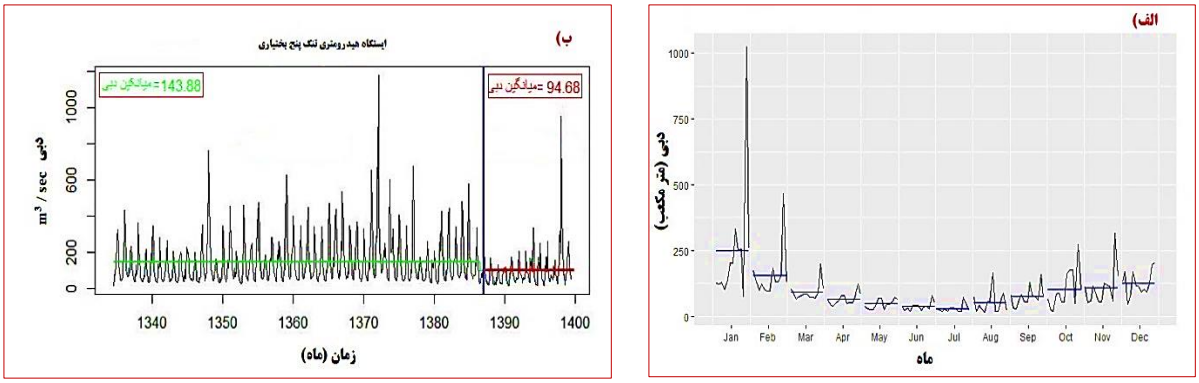
شکل ۳- مراحل یادگیری ماشین

از مراحل یادگیری ماشین می‌توان به تعیین مدل مناسب و مطالعه توصیفی داده‌ها اشاره کرد. باتوجه به شرایط و مقتضیات حاکم بر منطقه، تغییر در روند متغیرها همراه با تغییر در پارامترهای مدل صورت می‌گیرد؛ که این در پایداری پارامترهای مدل خلل ایجاد کرده و باعث تغییر رژیم آب شده است. تغییر رژیم آب در سری زمانی بدین معنی است که میانگین (و یا واریانس و یا هر دو) در دو قسمت قبل و بعد تغییر رژیم آب، از لحاظ آماری تفاوت معنی‌داری دارد. به‌عبارتی دو قسمت قبل و بعد سری زمانی، متعلق به دو جامعه متفاوت آماری می‌باشد. همان‌طورکه در شکل‌های (۴)، (۵) و (۶) مشاهده می‌شود این نقطهٔ تغییر رژیم آب از سال (۱۳۸۶) به‌دلیل انتقال آب بوده که باعث کاهش دبی ورودی به حوضهٔ آبریز دز شده است. البته عوامل مختلفی باعث تغییر الگوی ناگهانی سری‌های زمانی دبی نسبت به

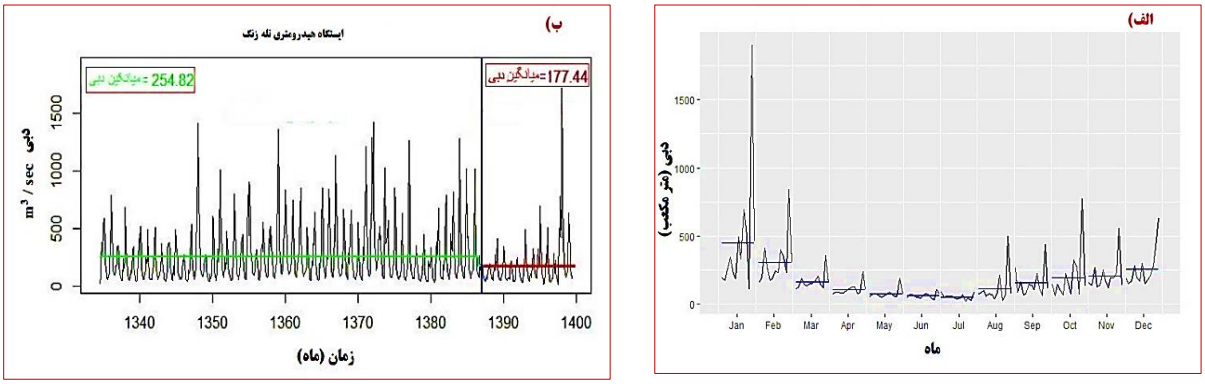
شرایط قبل از وقوع این تغییرات شده است؛ مثلاً احداث سد، وقوع آتش‌سوزی گسترده، تغییرات عمده در کاربری اراضی، تغییر اقلیم و غیره؛ بنابراین بررسی این تغییرات یک تحلیل پایه و ضروری است؛ و نیز از آنجایی‌که روش‌ها متناسب با ماهیت پدیده‌ها و پارامترهای مختلف، نحوه پراکنش و چگونگی روند تغییرات انتخاب می‌شوند؛ لذا این مقاله به‌جای استفاده کردن از تمام طول داده‌های سری زمانی از داده‌های سال (۱۳۸۶-۱۳۹۹) برای مدل‌سازی استفاده نموده است. قابل ذکر است ترکیب‌شدن انواع مدل‌های سری زمانی فضای حالت با الگوریتم داده‌کاوی بگینگ می‌تواند، یادگیری ماشین را نسبت به مدل باکس–جنکینز ارتقاء داده و باعث بهبود توانایی مدل در شبیه‌سازی جریان رودخانه شود. کاربرد این روش زمانی خود را بیشتر نشان می‌دهد که با حجم داده نسبتاً کمی مواجه باشیم.



شکل ٤- الف) میانگین سری زمانی ماهانه دبی در ایستگاه آب‌سنجی سپیددشت سزار از سال ۱۳۹۹-۱۳۸٦ (میانگین ماهانه با خط آبی مشخص شده است) ب) سری زمانی دبی در ایستگاه آب‌سنجی سپیددشت سزار از سال ۱۳۹۹-۱۳۳٤



شکل ٥- الف) میانگین سری زمانی ماهانه دبی در ایستگاه آب‌سنجی تنگ‌پنج بختیاری از سال ۱۳۹۹-۱۳۸٦ (میانگین ماهانه با خط آبی مشخص شده است) ب) سری زمانی دبی در ایستگاه آب‌سنجی تنگ‌پنج بختیاری از سال ۱۳۹۹-۱۳۳٤



شکل ٦- الف) میانگین سری زمانی ماهانه دبی در ایستگاه آب‌سنجی تله‌زنگ از سال ۱۳۹۹-۱۳۸٦ (میانگین ماهانه با خط آبی مشخص شده است) ب) سری زمانی دبی در ایستگاه آب‌سنجی تله‌زنگ از سال ۱۳۹۹-۱۳۳٤

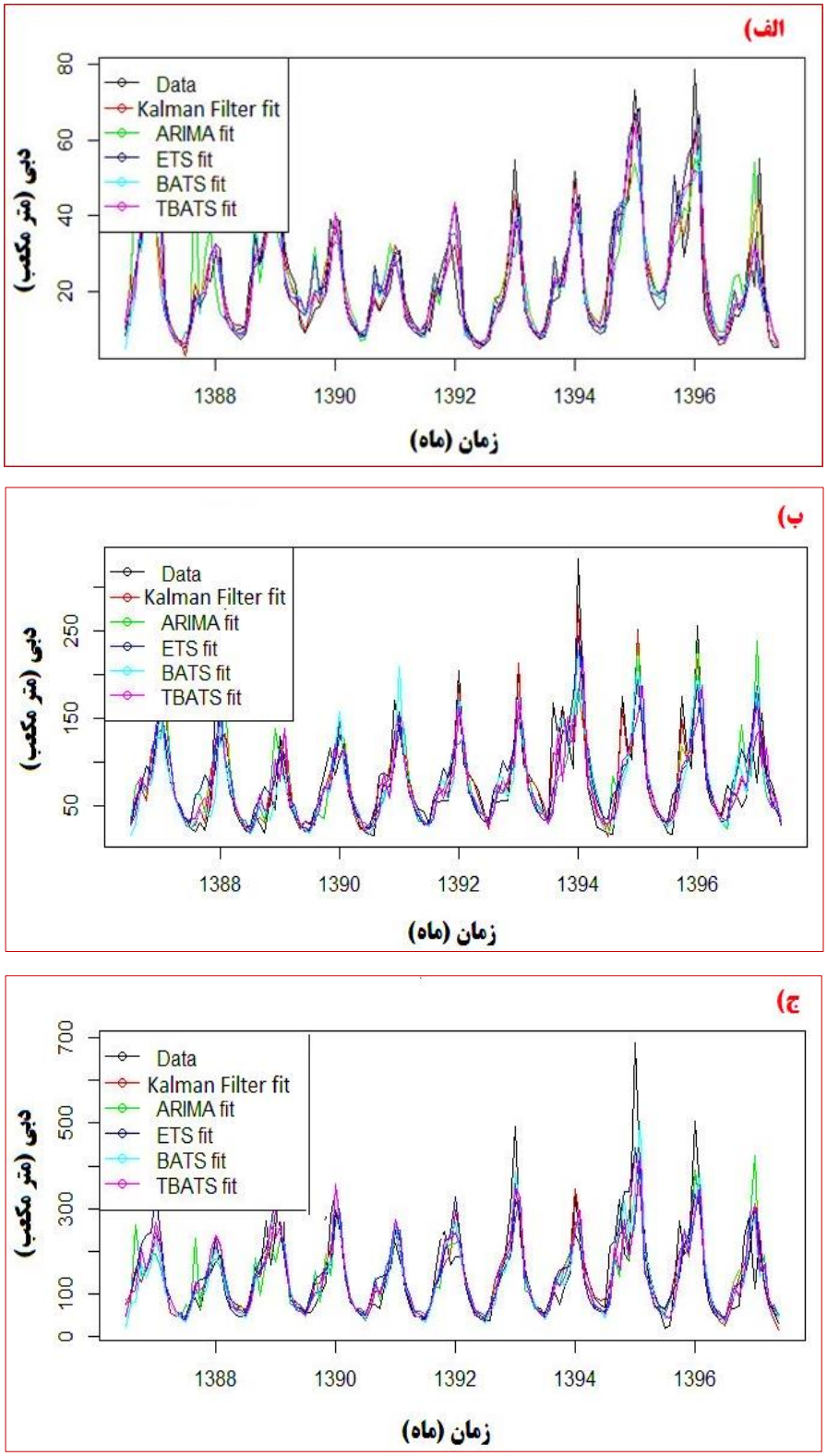
همچنین از شکل‌های (۴)،(۵) و (۶) مشاهده می‌شود که در طول سال‌های آبی، وضعیت پرآبی و کم آبی وجود دارد و همواره دبی در طول زمان در حال افزایش یا کاهش است؛ لذا داده‌ها دارای نوسان در میانگین هستند. یکی از بدترین تغییرات که می‌تواند در یک حوضه

آبریز اتفاق بیفتد و کلیه برنامه‌ریزی‌ها را مختل نماید تغییر در میانگین است. اما زمانی تغییر در واریانس در سری مشاهده می‌شود؛ که سری دچار بی‌نظمی‌های شدیدی شده است و پراکنش آن به‌معنای واقعی افزایش پیدا می‌کند. تغییرات در واریانس داده‌ها، به‌علت تغییرات شدید و افزایش در برداشت آب است که دبی را از حالت طبیعی خارج می‌نماید. اگر واریانس داده‌ها، تابعی از مقادیر گذشته باشد، نمی‌توان به‌راحتی واریانس داده‌ها را در طول زمان ثابت نگه داشت بنابراین از مدل‌های آریما نمی‌توان استفاده کرد؛ پس لازم است از مدل‌هایی که به فرض ثابت‌بودن واریانس نیاز ندارند، استفاده شود؛ که این مدل‌ها به خانواده گارچ[1] و آرچ[2] معروف هستند. اما رویکرد فضای حالت این مزیت را دارد که تغییرات در سری‌های زمانی را مدل نموده و به فرضیاتی هم‌چون ثابت‌بودن میانگین و واریانس نیازی ندارد.

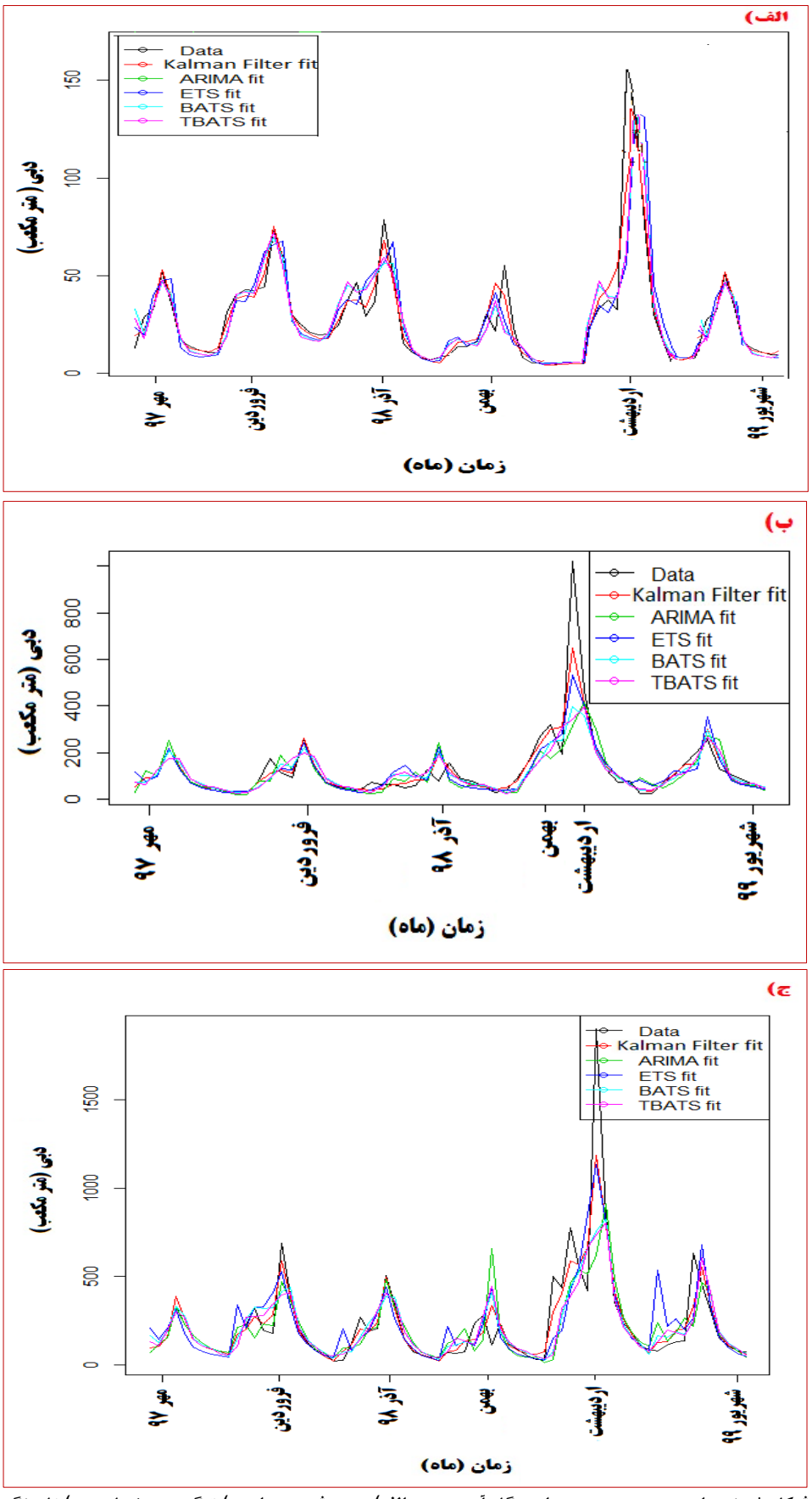
نتایج شبیه‌سازی دبی ماهانه

در این پژوهش، از چهار مدل بر پایه رویکرد فضای حالت و یک مدل بر پایه رویکرد کلاسیک (باکس- جنکینز) به‌منظور شبیه‌سازی دبی ماهانه سه ایستگاه آب‌سنجی حوضهٔ آبریز دز استفاده گردید. ابتدا پردازش اولیه داده‌ها انجام و سپس داده‌ها نرمال گردید تا از این طریق سرعت محاسبات افزایش یابد. سپس از ۸۵ درصد داده‌ها برای مرحله آموزش (واسنجی) و ۱۵ درصد برای تست مدل(صحت‌سنجی) در نظر گرفته شد؛ که نمودارهای شبیه‌سازی آن در شکل‌های (۷) و (۸) نشان داده شده است. با توجه به شکل (۸) که نتایج صحت‌سنجی مدل را در سه ایستگاه آب‌سنجی نشان می‌دهد مدل‌های فضای حالت با کمترین RMSE و بالاترین مقادیر R^2 در سه ایستگاه آب‌سنجی به‌عنوان بهترین مدل انتخاب شدند؛ هم‌چنین مدل‌های فضای حالت توانسته‌اند دبی‌های کوچکتر را بهتر شبیه‌سازی کنند؛ ولی در دبی‌های بالا نتوانسته‌اند انتظارات را برآورده کنند و مقادیر شبیه‌سازی پایین‌تر از مقادیر مشاهده‌ای بدست آمد. جدول (۲) نتایج کمی ارزیابی مدل‌های مورد استفاده را در دو مرحله واسنجی و صحت‌سنجی نشان می‌دهد. با توجه به جدول (۲)، مدل ترکیبی فیلتر کالمن که از زیر مجموعه رویکرد فضای حالت می‌باشد توانسته است نسبت به دیگر مدل‌ها در مرحله یادگیری و تست، قابلیت بهتری در شبیه‌سازی از خود نشان دهد. هم‌چنین بر اساس معیارهای ارزیابی خطا و دقت مدل در ایستگاه آب‌سنجی سپیددشت سزار مقدار R^2= ۰/۷۹ ، RMSE= ۳۹/۲۱ و MAE= ۱۷/۲۵ و در ایستگاه تنگ‌پنج بختیاری مقدارR^2=۰/۷۶ ، RMSE= ۵۷/۸۹ و MAE= ۲۴/۸۸ و در ایستگاه تله‌زنگ R^2=۰/۷۳، RMSE=۱۱۳/۴۱ و MAE= ۵۲/۹۱ به‌دست آمده است. هم‌چنین مشخص شد که رویکرد فضای حالت با استفاده از هر ساختاری در شبیه‌سازی متغیر هیدرولوژیکی دبی رودخانه تقریباً دقت بیشتری در مقایسه با رویکرد باکس- جنکینز دارد. در این راستا افراد ذیل به نتایج مشابهی دست یافتند. Boukharouba *etal*. (2016) عملکرد پیش‌بینی بین مدل‌های فضای حالت و باکس-جنکینز در دبی ماهانه و سالانه در شمال کشور الجزایر را مقایسه کردند نتایج آن‌ها نشان داد که برای هر دو مقیاس سالانه و ماهانه، مدل فضای حالت با استفاده از فیلتر کالمن در پیش‌بینی‌ها بهتر عمل می‌کند. Ahcen & OConnor (1994) از رویکرد فضای حالت و فیلتر کالمن برای پیش‌بینی دبی رودخانه استفاده نمودند و به این نتیجه رسیدند که در صورت وجود خطاهای اندازه‌گیری در سری زمانی، استفاده از روش فضای حالت و فیلتر کالمن به‌دلیل بهینه بدست آوردن پارامترهای مدل بهتر از مدل باکس-جنکینز عمل می‌کند. Li *etal*. (2017) فضای حالت دبی چشمه در یک حوضه کارست در جنوب غربی چین را پیش‌بینی نمودند نتایج آن‌ها نشان داد که مدل فضای حالت با استفاده از هر ترکیبی از متغیرها، از رگرسیون خطی کلاسیک، مدل شبکه عصبی مصنوعی و ماشین بردار پشتیبان در مدل‌سازی سری‌های زمانی برای چشمه کارست A عملکرد بهتری داشت. بنابراین، مدل‌سازی فضای حالت، ابزار مفیدی برای پیش‌بینی دبی چشمه در مناطق کارستی در جنوب غربی چین بود و این روش مدل‌سازی می‌تواند به محققان در دستیابی به نتایج دقیق در سایر مناطق کارستی کمک کند. هم‌چنین *etal*. Li (2017) تغییرات دبی رسوبی ماهانه و تخمین در یک حوضه کارست در جنوب غربی چین را مورد بررسی قرار دادند؛ نتایج آن‌ها نشان داد که روند کاهشی در دبی رسوبی در مقیاس ماهانه و سالانه مشاهده می‌شود. دبی آب و رسوب هر دو یک دوره یک ساله قابل توجهی داشتند که به این معنی است که تخلیه آب تأثیر قابل‌توجهی بر دبی رسوبی دارد. مدل فضای حالت با استفاده از دبی آب یک مدل مؤثر بود که ۹۹ درصد از تغییرات دبی رسوبی را به خود اختصاص داد. مدل فضای حالت به‌عنوان یک رویکرد مؤثر برای کمی‌کردن روابط زمانی بین دبی رسوبی و عوامل محرک آن در مناطق کارستی جنوب غربی چین توصیه شده است.

1 GARCH
2 ARCH



شکل ۷- نمودار واسنجی دبی ایستگاه آب‌سنجی الف)سپیددشت سزار ب) تنگ‌پنج بختیاری ج) تله‌زنگ (۱۳۹۷-۱۳۸۶)



شکل ۸- نمودار صحت‌سنجی دبی ایستگاه آب‌سنجی الف)سپیددشت سزار ب) تنگ‌پنج بختیاری ج) تله‌زنگ (۱۳۹۷-۱۳۹۹)

جدول ۲- مقایسه خطاهای واسنجی و صحت‌سنجی روش‌های مختلف با سه معیار MAE، RMSE و R^2 در ایستگاه‌های (سپیددشت سزار، تنگ‌پنج بختیاری و تله‌زنگ)

Discharge validation (1397-1399)			Discharge calibration (1386-1397)		Time series models	Station
R^2	RMSE (m^3/month)	MAE (m^3/month)	RMSE (m^3/month)	MAE (m^3/month)		
۰/۷۹	۳۹/۲۱	۱۷/۲۵	۱/۶۹	۱۸/۱۹	Kalman Filter	Sepiddasht Sezar
۰/۷	۴۸/۸۹	۲۵/۲۶	۲/۴۱	۷۸/۳۴	ETS	
۰/۶۸	۴۷/۷۵	۲۲/۰۹	۲/۳۷	۷۵/۲۶	SARIMA	
۰/۷۵	۶۹/۰۵	۳۹/۹۵	۲/۱۸	۶۳/۵۳	BATS	
۰/۷۷	۶۳/۵۹	۴۱/۹۳	۲/۲۴	۶۶/۴۵	TBATS	
۰/۷۶	۵۷/۸۹	۲۴/۸۸	۳/۳۱	۲۴۷/۲۳	Kalman Filter	Tangpanj Bakhtiyari
۰/۶۹	۷۹/۰۹	۴۱/۸۶	۴/۹۰	۱۰۷۸/۹۷	ETS	
۰/۵۸	۱۰۶/۲۶	۴۴/۳۲	۴/۸۰	۱۳۱۶/۱۹	SARIMA	
۰/۵۳	۱۰۱/۰۶	۴۰/۷۱	۴/۸۱	۱۱۲۴/۴۴	BATS	
۰/۵	۹۳/۰۵	۴۴/۳۸	۴/۷۸	۱۱۹۸/۴۸	TBATS	
۰/۷۳	۱۱۳/۴۱	۵۲/۹۱	۵/۶۸	۲۴۷۲/۶۹	Kalman Filter	Telezang
۰/۷	۱۶۵/۵۶	۹۴/۳۷	۶/۴۴	۳۸۸۳/۹۳	ETS	
۰/۵۸	۱۹۹/۷۴	۸۵/۴۹	۶/۱۳	۴۵۷۵/۸۹	SARIMA	
۰/۶۲	۱۸۲/۹۸	۸۴/۳۷	۵/۹۹	۳۵۶۶/۳۸	BATS	
۰/۵۱	۱۷۷/۹۴	۸۲/۱۰	۶/۰۵	۴۱۲۸/۸۴	TBATS	

نتیجه‌گیری

استفاده از مدل‌های سری‌زمانی با توجه به ماهیت تصادفی و غیرقطعی مسائل مدیریت منابع آب، به‌عنوان یکی از روش‌های توصیه شده در پیش‌بینی دبی رودخانه می‌باشد. لذا، در این مطالعه به‌منظور شبیه‌سازی دبی جریان رودخانه در سه ایستگاه حوضهٔ آبریز دز از چهار مدل سری زمانی فضای حالت و یک مدل کلاسیک (باکس–جنکینز) بر پایهٔ یادگیری ماشین استفاده گردید. نتایج نشان داد رویکرد فضای حالت جایگزین مناسبی نسبت به‌روش کلاسیک سری زمانی است؛ چرا که رویکرد فضای حالت می‌تواند در مرحله آموزش و تست بهتر از مدل باکس–جنکینز عمل کند. این نتایج با مطالعات Adubisi *etal.*(2019) و Husin *etal.* (2021) هم‌خوانی دارد. در بین همه مدل‌های فضای حالت، مدل سطح موضعی(فیلتر کالمن) از کارایی بالاتری برخوردار است. قابل ذکر است بیشترین اختلاف بین مقادیر مشاهداتی و شبیه‌سازی‌شده در اردیبهشت ۹۸می‌باشد که این به‌دلیل بارش‌های فراوان بوده و باعث افزایش چشمگیر دبی رودخانه شده است. با توجه به نتایج به‌دست آمده و با دلالت بر این نکته که بسیاری از فرآیندهای هیدرولوژیکی مانند کاهش و یا اوج دبی جریان بعد از سیلاب، معمولاً فرآیندی غیرخطی است؛ می‌توان پیشنهاد کرد جهت شناسایی روند، از الگوریتم‌های غیرخطی فضای حالت مانند:EKF و UKF استفاده گردد. تفاوت اساسی بین این دو الگوریتم این است که EKF معادلات غیرخطی را خطی کرده ولی UKF معادلات غیرخطی را مستقیماً مورد تحلیل و بررسی قرار می‌دهد.

"هیچ‌گونه تعارض منافع توسط نویسندگان وجود ندارد"

REFERENCES

Wu S.J., Lien H.C., Chang C.H., Shen J.C. (2012). Real-time correction of water stage forecast during rainstorm events using combination of forecast errors. *Stochastic Environmental Research and Risk* Assessment. *26*(4),519 531.

Nayak P.C., Sudheer K.P., Rangan D.M., Ramasastri K.S. (2004). A neurofuzzy computing technique for modeling hydrological time series. *Journal of Hydrology*, *291*,52–66.

El-Shafie, A., Taha, M. R., & Noureldin, A. (2007). A neuro-fuzzy model for inflow forecasting of the Nile river at Aswan high dam. *Water resources management*, *21*(3), 533-556.

Ngan, P. (1987). Kalman filter and its application to flow forecasting.

Sun, Y., Bao, W., Valk, K., Brauer, C. C., Sumihar, J., & Weerts, A. H. (2020). Improving forecast skill of

lowland hydrological models using ensemble Kalman filter and unscented Kalman filter. *Water Resources Research*, *56*(8), e2020WR027468.

Zeng, Q., Li, D., Huang, G., Xia, J., Wang, X., Zhang, Y., ... & Zhou, H. (2016). Time series analysis of temporal trends in the pertussis incidence in Mainland China from 2005 to 2016. *Scientific reports*, *6*(1), 1-8.

De Livera, A. M., Hyndman, R. J., & Snyder, R. D. (2011). Forecasting time series with complex seasonal patterns using exponential smoothing. *Journal of the American statistical association*, *106*(496), 1513-1527.

Lu, M., Lu, S., Liao, W., Lei, X., Yin, Z., & Wang, H. (2022). Research and application of parameter estimation method in hydrological model based on dual ensemble Kalman filter. *Hydrology Research*, *53*(1), 65-84.

Shen, Y., Wang, S., Zhang, B., & Zhu, J. (2022). Development of a stochastic hydrological modeling system for improving ensemble streamflow prediction. *Journal of Hydrology*, *608*, 127683.

Stefenon, S. F., Ribeiro, M. H. D. M., Nied, A., Yow, K. C., Mariani, V. C., dos Santos Coelho, L., & Seman, L. O. (2022). Time series forecasting using ensemble learning methods for emergency prevention in hydroelectric power plants with dam. *Electric Power Systems Research*, *202*, 107584.

Dong, X., Yu, Z., Cao, W., Shi, Y., & Ma, Q. (2020). A survey on ensemble learning. *Frontiers of Computer Science*, *14*(2), 241-258.

Fletcher, R. (2013). *Practical methods of optimization*. John Wiley & Sons.

Pintelas, P., & Livieris, I. E. (2020). Special issue on ensemble learning and applications. *Algorithms*, *13*(6), 140.

Liu, B. (2009). Some research problems in uncertainty theory. *Journal of Uncertain systems*, *3*(1), 3-10.

Breiman, L. (1996). Bagging predictors. *Machine learning*, *24*(2), 123-140.

Huang, F., Xie, G., & Xiao, R. (2009, November). Research on ensemble learning. In *2009 International Conference on Artificial Intelligence and Computational Intelligence* (Vol. 3, pp. 249-252). IEEE.

Dantas, T. M., & Oliveira, F. L. C. (2018). Improving time series forecasting: An approach combining bootstrap

aggregation, clusters and exponential smoothing. *International Journal of Forecasting*, *34*(4), 748-761.

Bühlmann, P., & Yu, B. (2002). Analyzing bagging. *The annals of Statistics*, *30*(4), 927-961.

Bühlmann, P. (2012). Bagging, boosting and ensemble methods. In *Handbook of computational statistics* (pp. 985-1022). Springer, Berlin, Heidelberg.

Box, G. E. P., & Jenkins, G. M. (1970). 1970: Time series analysis, forecasting and control. San Francisco: Holden-Day.

Makridakis, S., & Hibon, M. (1997). ARMA models and the Box–Jenkins methodology. *Journal of forecasting*, *16*(3), 147-163.

Shumway, R. H., & Stoffer, D. S. (1982). An approach to time series smoothing and forecasting using the EM algorithm. Journal of time series analysis, 3(4), 253-264.

Shumway, R. H. (1988). Applied statistical time series analysis. *Prentice Hall Series in Statistics*.

Kalman R.E. (1960). A new approach to linear filtering and prediction problems. *Journal of Basic Engineering Transactions*, *82*(1), 35-45.

Chan, N. H., & Palma, W. (1998). State space modeling of long-memory processes. *The Annals of Statistics*, *26*(2), 719-740.

Sharifi, M., Mohammadzadeh Shobegar, A., Radmanesh, F., & Mansouri, B. (2022). Evaluation of structural approaches to state space compared to classical in predicting precipitation time series (Dez catchment). *Water and Irrigation Management*, *12*(1), 1-13.

Durbin, J., & Koopman, S. J. (2012). *Time series analysis by state space methods* (Vol. 38). OUP Oxford.

Harvey, A. C. (1990). Forecasting, structural time series models and the Kalman filter.

Shamshad, B., Khan, M. Z., & Omar, Z. (2019). Modeling and forecasting weather parameters using ANN-MLP, ARIMA and ETS model: a case study for Lahore, Pakistan. *Journal of Applied Statistics*, *5*, 388.

Jain, G., & Mallick, B. (2017). A study of time series models ARIMA and ETS. *Available at SSRN 2898968*.

Alakkari, K. M., Mishra, P., Rawat, D., Abotaleb, M., & Ghazi, A. M. (2022). Using ETS State Space Model for forecasting on third wave on COVID19 in India.

Abd Rahman, N., Yusop, Z., Åžen, Z., Taher, S., & Kane, I. L. (2017). Mitigation of time series approach on climate change adaptation on rainfall of Wadi Al-Aqiq, Madinah, Saudi Arabia. *Jurnal Teknologi*, *79*(5).

Hyndman, R., Koehler, A. B., Ord, J. K., & Snyder, R. D. (2008). *Forecasting with exponential smoothing: the state space approach*. Springer Science & Business Media.

Gos, M., Krzyszczak, J., Baranowski, P., Murat, M., & Malinowska, I. (2020). Combined TBATS and SVM

model of minimum and maximum air temperatures applied to wheat yield prediction at different locations in Europe. *Agricultural and Forest Meteorology*, *281*, 107827.

De Livera, A. M. (2010). Automatic forecasting with a modified exponential smoothing state space framework. Monash Econometrics and Business Statistics Working Papers, 10(10), 6.

Badr, A., Makarovskikh, T., Mishra, P., Abotaleb, M., Al Khatib, A. M. G., Karakaya, K., ... & Attal, E. (2021). Modelling and forecasting of web traffic using Holt's linear, bats and TBATS models. *J. Math. Comput. Sci.*, *11*(4), 3887-3915.

Naim, I., Mahara, T., & Idrisi, A. R. (2018). Effective short-term forecasting for daily time series with complex seasonal patterns. *Procedia computer science*, *132*, 1832-1841.

Zachary, H., Asghar, Z., & Oroza, C. A. (2020). Multi-step Weekly Average Forecasting of Reservoir Storage Volume Using Deep Learning. *Earth and Space Science Open Archive ESSOAr*.

Huang, K., Zhang, J., & Song, Y. (2022, May). Application of machine learning models based on ANN and GA coupling algorithms in hydrological runoff simulation. In *International Conference on Electronic Information Engineering, Big Data, and Computer Technology (EIBDCT 2022)* (Vol. 12256, pp. 373-380).

Yeh, W.W.-G.; Yoon, Y.S. (1981). Aquifer parameter identification with optimum dimension
in parameterization. Water Resources Research. 17, 664–672.

Kotsiantis, S. B., Zaharakis, I. D., & Pintelas, P. E. (2006). Machine learning: a review of classification and combining techniques. Artificial Intelligence Review, 26(3), 159-190.

Ribeiro, M. H. D. M., & dos Santos Coelho, L. (2020). Ensemble approach based on bagging, boosting and stacking for short-term prediction in agribusiness time series. *Applied Soft Computing*, *86*, 105837.

Gastinger, J., Nicolas, S., Stepić, D., Schmidt, M., & Schülke, A. (2021, July). A study on ensemble learning for time series forecasting and the need for meta-learning. In *2021 International Joint Conference on Neural Networks (IJCNN)* (pp. 1-8). IEEE.

Murray, S. A. (2018). The importance of ensemble techniques for operational space weather forecasting. *Space Weather*, *16*(7), 777-783.

Seising, R., & Allende-Cid, H. (Eds.). (2017). *Claudio Moraga: A Passion for Multi-Valued Logic and Soft Computing*. Springer International Publishing.

Wideroos, K. (2021). Improving time series forecasting accuracy by ensemble methods and time series features (Master's thesis).

Li, Y., Liang, Z., Hu, Y., Li, B., Xu, B., & Wang, D. (2020). A multi-model integration method for monthly streamflow prediction: modified stacking ensemble strategy. *Journal of Hydroinformatics*, *22*(2), 310-326.

Rajesh, M., & Rehana, S. (2021). Prediction of river water temperature using machine learning algorithms: a tropical river system of India. *Journal of Hydroinformatics*, *23*(3), 605-626.

Wang, W. C., Chau, K. W., Xu, D. M., & 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.

Ardabili, S., Mosavi, A., & Várkonyi-Kóczy, A. R. (2019, September). Advances in machine learning modeling reviewing hybrid and ensemble methods. In *International Conference on Global Research and Education* (pp. 215-227). Springer, Cham.

Boukharouba, K., & Kettab, A. (2016). Comparison of prediction performances between Box–Jenkins and Kalman filter models—Case of annual and monthly sreamflows in Algeria. Desalination and Water Treatment, 57(36), 17095-17103.

Ahsan, M., & O'Connor, K. M. (1994). A reappraisal of the Kalman filtering technique, as applied in river flow forecasting. *Journal of Hydrology*, *161*(1-4), 197-226.

Li, Z., Xu, X., Liu, M., Li, X., Zhang, R., Wang, K., & Xu, C. (2017). State-space prediction of spring discharge in a karst catchment in southwest China. *Journal of Hydrology*, *549*, 264-276.

Li, Z., Xu, X., Xu, C., Liu, M., Wang, K., & Yi, R. (2017). Monthly sediment discharge changes and estimates in a typical karst catchment of southwest China. *Journal of Hydrology*, *555*, 95-107.

Adubisi, O. D., David, I. J., Eka, O., & Uduma, A. E. (2019). State space and Box-Jenkins approaches: a comparison of models prediction performance in finance. *International Journal of Data Science*, *4*(3), 181-195.

Husin, W. Z. W., Afdzal, A. S., Azmi, N. L. H., & Hamadi, S. A. T. S. (2021, November). Box-Jenkins and State Space Model in Forecasting Malaysia Road Accident Cases. In *Journal of Physics: Conference Series* (Vol. 2084, No. 1, p. 012005). IOP Publishing.