Archive of SID.ir

تحقیقات منابع آب ایران
Iran-Water Resources Research

سال بیستم، شماره ۱، بهار ۱۴۰۳
Volume 20, No. 1, Spring 2024 (IR-WRR)
۱۸۱-۱۹۶

تخمین داده‌های اندازه‌گیری نشده دبی ایستگاه آب‌سنجی با استفاده از مدل‌های ترکیبی رگرسیونی ماشین‌های بردار پشتیبان و راهکار فیلتر کالمن تجمیعی (مطالعه موردی: بالادست حوضه زاینده‌رود)

علیرضا کاظم‌زاده ورزنه[۱]، مهدی فاضلی[۲*] و حسین منتصری[۳]

چکیده

به‌طورمعمول داده‌های ایستگاه‌های هواشناسی جهت توسعه و صحت‌سنجی عملکرد مدل‌های هیدرولوژیکی به کار می‌روند. عدم ثبت و اندازه‌گیری داده‌های هیدرولوژی و هواشناسی موجب کاهش دقت مدل‌های هیدرولوژی و یا عدم امکان ساخت و توسعه آن‌ها می‌شود. در این مطالعه بازسازی خلأ آماری داده‌های آب‌سنجی ایستگاه چلگرد واقع در بالادست زیر حوضه قلعه‌شاهرخ-چلگرد از حوضه آبریز زاینده‌رود مورد بررسی قرار گرفت. ایستگاه یادشده میزان آورد آب تونل اول کوهرنگ را اندازه‌گیری می‌کند. برای تخمین داده‌های ثبت نشده از مدل رگرسیونی ماشین بردار پشتیبان و به‌منظور بهبود عملکرد مدل موردنظر و کاهش خطای شبیه‌سازی، از راهکار داده‌گواری فیلتر کالمن تجمیعی (EnKF) استفاده شد. شاخص‌های ضریب همبستگی، معیار ریشه میانگین مجذور خطا و ضریب نش-ساتکلیف به‌منظور سنجش عملکرد مدل رگرسیونی جهت تخمین داده‌های دبی جریان استفاده شد. مقادیر شاخص‌های مذکور به ترتیب برابر ۰/۸۳، ۳/۴۲ و ۰/۷۱ برای مرحله آموزش و مقادیر ۰/۷۰، ۲۰/۳۸ و ۰/۲۵ برای مرحله آزمون بدست آمد که با به‌کارگیری EnKF عملکرد مدل رگرسیونی ارتقاء یافت و نتایج قابل‌قبولی با بازتولید داده‌ها بدست آمد. برای اصلاح نتایج EnKF از داده‌های ایستگاه قلعه‌شاهرخ واقع در خروجی حوضه آبریز به عنوان ایستگاه شاهد با کمک یک مدل رگرسیونی SVR دوم استفاده شد. برای مدل رگرسیونی اخیر، مقادیر شاخص‌های مذکور به ترتیب برابر ۰/۹۶، ۵/۲ و ۰/۸۱ برای مرحله آموزش و ۰/۷۶، ۶/۶ و ۰/۶۶ برای مرحله آزمون به‌دست آمد.

کلمات کلیدی: داده‌های ناموجود دبی جریان سطحی، بازسازی داده‌های ثبت نشده، ماشین‌های بردار پشتیبان، مدل رگرسیونی، داده‌گواری.

تاریخ دریافت مقاله: ۱۴۰۲/۱۰/۱۵

تاریخ پذیرش مقاله: ۱۴۰۳/۲/۱۲

۱- دانش‌آموخته کارشناسی ارشد، گروه مهندسی عمران، دانشگاه یاسوج، یاسوج، ایران.
۲- استادیار گروه مهندسی عمران، دانشگاه یاسوج، یاسوج، ایران.
۳- دانشیار گروه مهندسی عمران، دانشگاه یاسوج، یاسوج، ایران.
*- نویسنده مسئول

بحث و مناظره (Discussion) در مورد این مقاله تا پایان تابستان ۱۴۰۳ امکانپذیر است.

Estimation Missing Stream Flow Data of Hydrometric Gauge Using Support Vector Regression and Ensemble Kalman Filter (EnKF) Technique (Case Study: Upstream Zayandehrud Basin)

A.R Kazemzadeh Varzaneh [1], M. Fazeli[2], and H. Montaseri[3]*

Abstract

Measuring and recording climate data of gauges are usually used to develop and calibrate hydrological models. Missing hydrological and climate data can decrease models' accuracy or impede developing the models. In this study, remaking the missing data of Chelgerd hydrometric gauge located at the upstream of Ghaleshahrokh-Chelgerd sub basin as part of Zayandehrud Basin was surveyed. It measures the discharge from the first Koohrang tunnel inflow. In order to estimate the missing data, regression support vector machine model was employed and to improve the model performance, Ensemble Kalman Filter (EnKF) was used as data assimilation technique. For evaluating the regression model performance, R, RMSE and Nash-Sutcliff citeria were implemented. The results showed values of 0.83, 3.42, 0.7 for training part and the values of 0.70, 20.38 and 0.25 for test part of the model, for R, RMSE and NS respectively. By using EnKF, the performance of the regression model has been improved and acceptable results were obtained. To modify the EnKF results, the data of Ghaleshahrokh station as reference station located at basin outlet was used alongside a second SVR model. The values of R, RMSE and NS were respectively 0.96, 5.2 and 0.81 for training and 0.76, 6.6 and 0.66 for test stage.

Keywords: Non-Measuring Inflow Discharge, Estimating Missing Data, Support Vector Machine, Regression Model, Data Assimilation.
Received: January 5, 2024
Accepted: May 1, 2024

1- M.Sc. Graduated, Yasouj University, Yasouj, Iran. Email: kazemzadeh711@gmail.com
2- Assistant Professor, Department of Civil Engineering, Yasouj University, Iran. Email: fazeli@yu.ac.ir
3- Associate Professor, Department of Civil Engineering, Yasouj University, Iran. Email: hmontaseri@yu.ac.ir
*- Corresponding Author
Doi:10.22034/IWRR.2024.382843.2735

This work is licensed under a Creative Commons Attribution 4.0 International License

Archive of SID.ir

Archive of SID.ir

۱- مقدمه

اهمیت داده‌های اندازه‌گیری زمینی برای توسعه یک مدل هیدرولوژیکی اجتناب‌ناپذیر است. به طور معمول داده‌های هواشناسی از جمله بارش، دما، آب‌سنجی و پارامترهای هیدرولوژی نظیر سرعت و جهت باد، رطوبت و تابش خورشید برای توسعه و صحت‌سنجی عملکرد مدل‌های هیدرولوژیکی به کار می‌روند. استفاده از ابزار مناسب با دقت بالا جهت اندازه‌گیری پارامترهای موردنیاز از اهمیت بالایی برخوردار است. به طور معمول، عدم ثبت داده‌های اندازه‌گیری شده، عدم کفایت تعداد ایستگاه‌های اندازه‌گیری و خطای ابزارها از جمله مواردی است که سبب می‌شود با بحران و یا چالش کمبود داده‌های مشاهداتی مواجه شویم؛ در نتیجه صحت‌سنجی و توسعه مدل‌های شبیه‌ساز هیدرولوژیکی و یا تحلیل داده‌ها با مشکل روبرو می‌شود. بنابراین خطاهای انسانی و ابزاری یکی از چالش‌های مهم در ثبت بانک اطلاعات مفید هیدرولوژیکی است که می‌تواند به عدم ثبت آمار یا تشخیص داده‌های پرت، حذف آن‌ها و معضل داده‌های ثبت نشده منجر شود.

به‌منظور بررسی و تحلیل آماری داده‌ها و به‌کارگیری در مدل‌سازی‌های هیدرولوژیکی لازم است تا داده‌های ثبت نشده یا اندازه‌گیری نشده بازسازی شوند؛ برای این منظور روش‌های مختلفی استفاده می‌شود. اغلب این روش‌ها بر پایه و اساس داده‌های مشاهداتی ایستگاه‌های مجاور است. در برخی از مطالعات بر اساس داده‌های ایستگاه‌های مجاور، داده‌های ثبت نشده جریان سطحی بازسازی شده است (Hirsch, 1979; Wallis et al., 1991). از دیگر روش‌های بازسازی داده‌های هیدرولوژیکی و هواشناسی، روش‌های بر پایه روابط رگرسیونی را می‌توان نام برد. برای نمونه، ارائه تابع رگرسیونی جهت تخمین داده‌های بارندگی ماهانه یک حوضه آبریز بر اساس مشاهدات ۳۰ سال داده باران‌سنجی از ۵ ایستگاه مجاور (Mcculloch and Booth, 1970)، استفاده از پنج روش رگرسیونی جهت بازسازی داده جریان سطحی در یکی از زیر حوضه‌های آبریز هند (Raman et al., 1995)، ارزیابی روش‌های بازتولید داده جهت بازسازی داده‌های ثبت نشده سطح ایستابی چاه‌های مشاهداتی حوضه آبریز کازرون واقع در ایران و عملکرد ارزیابی یک مدل رگرسیونی در مقایسه با روش‌های سنتی نظیر میانگین‌گیری حسابی و درون‌یابی Spline (Asgharinia and Petroselli, 2020) از سوابق مطالعات پیشین در این زمینه است. در دهه‌های اخیر، کاربرد مدل‌های داده محور در زمینه تولید داده و یا بازسازی داده‌های ثبت نشده افزایش یافته است. در ادامه چند نمونه مرور می‌شود. بازتولید داده‌های ثبت نشده بارش ایستگاه‌های حوضه آبریزی واقع در آتلانتیک شمالی آمریکا با استفاده از روش پس انتشار شبکه عصبی (Kuligowski and Barros, 1998)، استفاده از مدل شبکه عصبی برای بازسازی داده‌های جریان سطحی ۵ رودخانه (Khalil et al., 2001)، به‌کارگیری مدل داده محور شبکه عصبی بازگشتی جهت بازسازی داده‌های ناموجود جریان سطحی در سه ایستگاه آب‌سنجی و مقایسه نتایج به‌دست‌آمده با مدل خطوط رگرسیون تطبیقی چند متغیره بهینه (Coulibaly and Baldwin, 2005)، بازتولید داده‌های دبی ۱۰ ایستگاه آب‌سنجی واقع در نقاط مختلف ایران با استفاده از مدل‌های شبکه عصبی و نرو فازی و مقایسه نتایج بدست آمده با روش‌های قدیمی نظیر روش نسبت نرمال و همبستگی (Dastorani et al., 2010)، استفاده از مدل رگرسیونی ماشین‌های بردار پشتیبان جهت بازتولید و شبیه‌سازی دبی جریان سطحی برای رخدادهای مختلف هیدرولوژیکی از جمله شبیه‌سازی دبی اوج سیلاب، ذوب برف، رواناب ناشی از بارش‌های سنگین رگبارها و جریان‌هایی با دبی کم (Langhammer and Cesak, 2016) معرفی یک روش داده‌کاوی جدید جهت بازسازی داده‌های دبی جریان سطحی حوضه‌های آبریز ایالات واشنگتن و آیداهو آمریکا (Petty and Dhingra, 2018)، توسعه یک روش داده محور (روش انتساب مبتنی بر فرکانس)، برای بازسازی ۳۸۵ داده ثبت نشده و ۵۵۴۳ داده مشاهداتی ایستگاه‌های آب‌سنجی در حوضه آبریز آنتالیا واقع در ترکیه (Dikbas and Yasar, 2020) و استفاده از دو مدل داده محور رگرسیونی بردار پشتیبان و روند رگرسیونی گاوسی جهت بازسازی داده‌های ثبت نشده ایستگاه‌های بارش واقع در ترکیه و مقایسه نتایج حاصل از دو مدل با روش‌های حسابی درختی نظیر روش جنگل تصادفی (Sattari et al., 2020).

در اغلب مواقع، کمبود بانک مناسب اطلاعاتی از داده‌ها، سبب می‌شود عملکرد مدل‌های شبیه‌ساز با عدم قطعیت روبرو شود. به‌خصوص برای علومی نظیر هیدرولوژی که با محدودیت داده و اطلاعات مواجه است. فنون داده‌گواری[1] یکی از روش‌های هوشمندی است که برای پردازش و بهنگام سازی داده‌ها به کار می‌رود. کاربرد داده‌گواری در زمان‌های گذشته بیشتر در زمینه هواشناسی و اقیانوس‌شناسی محدود بوده اما در سال‌های اخیر این فنون به‌طور گسترده در کلیه زمینه‌های هیدرولوژی نظیر سنجش از دور بر روی اطلاعات سطحی زمین به کار برده شده و بهنگام سازی داده‌های مربوط به رطوبت خاک، پوشش برف و دما مورد مطالعه قرار گرفته است (Crow and Wood, 2003; Andreadis and Lettenmaier, 2006; Bosilovich et al., 2007, Kumar et al., 2008). بهبود عملکرد مدل داده محور ماشین‌های بردار پشتیبان به‌منظور پیش‌بینی روزانه رطوبت خاک که خروجی مدل توسط راهکار فیلتر کالمن تجمیعی و داده‌های مشاهداتی

Archive of SID.ir

Archive of SID.ir

(به عنوان داده‌های اندازه‌گیری شده) بهنگام‌سازی می‌شود. داده‌های بهنگام سازی شده به عنوان ورودی مدل ماشین بردار پشتیبان[۲] برای گام زمانی آتی درنظر گرفته می‌شود که این امر سبب اصلاح مدل آموزشی SVM در مراحل بعدی پیش‌بینی می‌شود (Kashif Gill et al., 2007; Liu et al., 2010)؛ استفاده از مدل‌های رگرسیونی ماشین‌های بردار پشتیبان جهت شبیه‌سازی مدل بارش–رواناب برای حوضه آبریزی در کشور چین و تلفیق آن با روش فیلتر کالمن تجمیعی برای کاهش خطای تولیدشده در حین شبیه‌سازی بطوریکه با سنجش معیار ارزیابی خطای RMSE[3]، میزان مقدار خطا در محاسبه دبی جریان از ۳۰/۶ مترمکعب بر ثانیه (بدون تلفیق) به مقدار ۱۳/۵ مترمکعب بر ثانیه (با تلفیق) کاهش یافت (Li et al., 2014)، بهبود عملکرد یک مدل داده محور رگرسیونی ماشین‌های بردار پشتیبان با کمک راهکار اصلاح یافته داده‌گواری (تلفیق راهکار فیلتر کالمن تجمیعی و یک مدل بهینه‌سازی جهت بهبود عملکرد مدل داده محور) در شبیه‌سازی فرایند بارش–رواناب که در یک مدل بهینه‌سازی، بهنگام‌سازی خروجی مدل تا جایی ادامه می‌یابد که خطای شبیه‌سازی از خطای شبیه‌سازی تکرار قبل کمتر شده و سپس مدل برای گام زمانی آتی شبیه‌سازی انجام می‌دهد (Mehrparvar and Asghari, 2018)، کاهش خطای شبیه‌سازی مدل بارش–رواناب SWAT با استفاده از فیلتر کالمن تجمیعی که بهنگام‌سازی داده‌های در این مطالعه شامل داده‌های ورودی هواشناسی و پارامترهای مدل بوده است (Mehrparvar et al., 2019). استفاده از مدل ماشین‌های بردار پشتیبان به‌منظور پیش‌بینی رطوبت در لایه‌های مختلف خاک اراضی کشاورزی و استخدام راهکار داده‌گواری فیلتر کالمن تجمیعی جهت ارتقای عملکرد مدل داده محور بطوریکه با محاسبه شاخص‌های ارزیابی خطا، مقدار ضریب همبستگی R به بیش از ۰/۸ بدست آمده و مقدار کاهش خطا با شاخص RMSE به میزان ۶۰ درصد نسبت به حالت عدم بکارگیری راهکار مورد نظر بوده است (Zhu et al., 2021) و بررسی اثر داده‌های بهنگام‌سازی‌شده ماهواره‌ای رطوبت خاک بر داده‌های پیش‌بینی جریان سطحی مدل توزیعی HLM با به‌کارگیری سه راهبرد متعدد از جمله راهکار فیلتر کالمن تجمیعی جهت به‌روزرسانی داده‌های پیش‌بینی‌شده رطوبت خاک (Jadidoleslam et al., 2021) از جمله شواهدی است که ادعای بهبود قدرت پیش‌بینی و پردازش مدل‌های داده محور در اثر به‌کارگیری فنون داده‌گواری را تقویت می‌کند.

با توجه به بررسی‌ها و ارزیابی مطالعات گذشته پیرامون بازسازی داده‌های ثبت نشده در زمینه علوم هیدرولوژی، اغلب مطالعات به استفاده از مدل رگرسیونی یا داده محور در جهت بازسازی داده‌ها اکتفا می‌کنند و از مقایسه نتایج بدست آمده با داده‌های مشاهداتی، عملکرد این مدل‌ها را مناسب ارزیابی می‌کنند. از طرفی استفاده از ابزارها و روش‌های داده‌گواری نظیر فیلتر کالمن برای بهنگام‌سازی و ارتقاء عملکرد مدل‌های شبیه‌ساز با در دسترس بودن داده‌های موجود بوده است و خلاء مطالعاتی در کاربرد فنون داده‌گواری در بازسازی داده‌های ثبت نشده و یا اندازه‌گیری نشده وجود دارد. به منظور بررسی میزان توانایی این فنون در بازسازی داده‌های بدون اندازه‌گیری، در این مطالعه از فن داده‌گواری جهت بهنگام‌سازی خروجی مدل و ارتقاء مدل شبیه‌ساز استفاده شده است. برای بررسی میزان توانایی این مدل‌ها در بازسازی داده‌های ثبت‌نشده، لازم است تا حد امکان تمهیداتی در جهت بهبود پردازش این‌گونه مدل‌ها به کار گرفته شود. بدین ترتیب در این مطالعه، با هدف بازسازی داده‌های ثبت نشده ایستگاه آب‌سنجی چلگرد واقع در زیر حوضه آبریز چلگرد–قلعه شاهرخ (از زیر حوضه‌های بالادست حوضه آبریز بزرگ زاینده‌رود) با استفاده از مدل رگرسیونی ماشین‌های بردار پشتیبان، داده‌های دبی جریان سطحی ایستگاه موردنظر بازتولید می‌شود. برای حصول اطمینان از نتایج مدل ساخته‌شده و کاربرد آن در بازسازی داده‌های آماری هیدرولوژی، از راهکار داده‌گواری در جهت بهبود فرایند شبیه‌سازی استفاده می‌شود تا داده‌های مورد قبول و قابل اعتمادی ارائه شود.

۲- روش تحقیق

با توجه به موضوع این پژوهش در خصوص تولید داده‌های ثبت نشده، از مدل رگرسیونی بردار پشتیبان جهت تولید داده استفاده شد و داده‌های بازسازی‌شده توسط راهکار فیلتر کالمن تجمیعی اصلاح شد. پارامتر موردنظر جهت بازسازی، دبی جریان سطحی است. بر اساس داده‌های اندازه‌گیری موجود، دبی جریان سطحی توسط مدل‌های تجربی و مفهومی قابل برآورد است که در این تحقیق از مدل رگرسیونی بردار پشتیبان[۴] جهت برآورد دبی جریان سطحی در محل ایستگاه‌های آب‌سنجی که دارای خلاء آماری است، استفاده شد و به‌منظور کاهش خطای شبیه‌سازی، راهکار فیلتر کالمن تجمیعی[۵] جهت بهنگام‌سازی و اصلاح نتایج شبیه‌سازی به‌کار رفت.

۲-۱- مدل رگرسیونی بردار پشتیبان

هدف از مدل رگرسیون بردار پشتیبان که با نام الگوریتم ε -SVR نیز معرفی می‌شود (Vapnik, 1995)، پیدا کردن یک تابع هموار است که با حذف داده‌های پرت و حداقل کردن خطای ممکن، یک الگوی آموزش‌دیده مناسب ارائه دهد. فرض شود مجموعه جفت داده‌های ورودی و خروجی (x,y) مفروض است. در مدل‌سازی رگرسیونی

Archive of SID.ir

Archive of SID.ir

میزان انحراف ε از داده‌های خروجی به کمترین مقدار خود برای ارائه تابع $y_i = f(x_i)$ درنظر گرفته می‌شود. مطابق با شکل ۱، ε مقداری است که اجازه می‌دهد تا خروجی‌های تابع از داده‌های خروجی واقعی انحراف یابند و در نتیجه تابع هموارساز حاصل شود. بنابراین مقادیر خارج از همسایگی ε، به‌عنوان خطا فرض می‌شود که مدل‌سازی بر اساس حداقل کردن خطاهای بیشتر توسعه می‌یابد و بدین صورت تابع خطای ε-insensitive به شکل رابطه (۱) و شکل خطی توابع رگرسیونی به صورت رابطه (۲) تعریف می‌شود (Smola and Scholkopf, 2004):

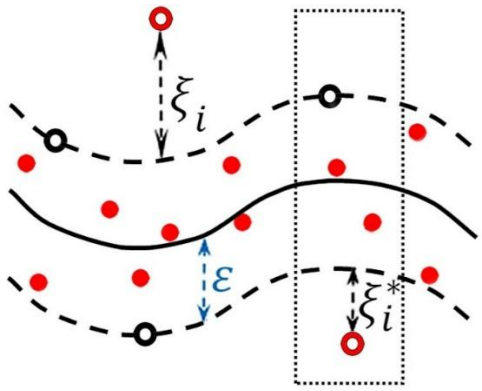


Fig. 1- Scattering of data in the neighborhood of ε limit (Noori et al., 2011)

شکل ۱- پراکندگی داده‌ها در همسایگی حدی ε مدل SVM (Noori et al., 2011)

$$|\xi|_\varepsilon = |y - f(x)|_\varepsilon = \begin{cases} 0 & \text{if } |y - f(x)| \le \varepsilon \\ |y - f(x)| - \varepsilon & \text{otherwise} \end{cases} \quad (۱)$$

$$f(x) = w.x + b \quad (۲)$$

w بردار وزنی است که به‌صورت ضرب داخلی در بردار ورودی‌های x ضرب می‌شود. هسته اصلی ماشین‌های بردار پشتیبان، یک مدل بهینه‌سازی است که در این مدل با هدف حداقل کردن خطاها، توابع رگرسیونی بدست می‌آیند. مدل بهینه‌سازی در جهت حداقل کردن خطاها در مدل‌های ماشین بردار پشتیبان به شکل روابط (۳) است (Vapnik, 1995). در روابط نامبرده با هدف کاهش بردار وزن و درنظر گرفتن ضریب جریمه برای مقادیر تخطی از ξ_i و ξ_i^*، توابع هموارساز تولیدشده به‌عنوان روابط رگرسیونی جهت شبیه‌سازی به دست می‌آیند.

$$\text{Minimize} \frac{1}{2}||w||^2 + C\sum_{i=1}^{n}(\xi_i + \xi_i^*) \quad (۳)$$

subject to:

$$f(x_i) - y_i \le \varepsilon + \xi_i$$

$$y_i - f(x_i) \le \varepsilon + \xi_i^*$$

$$\xi_i, \xi_i^* \ge 0, i = 1,2,\dots,n$$

ثابت جریمه C به عنوان پارامتر ظرفیت تابع نامیده شده که بیانگر سازشی بین پیچیدگی تابع هدف و خطای قابل‌قبول است. y نیز بردار داده‌های مشاهداتی بوده که برای هر x_i ورودی y_i نظیر آن وجود دارد. به طور کلی، مدل‌های رگرسیونی بردار پشتیبان دارای دو پارامتر اصلی و کلی شامل پارامتر همسایگی ε و ثابت جریمه C هستند. معمولاً در زمینه علوم هیدرولوژی، روابط خطی همبستگی مناسبی بین پارامترهای هیدرولوژی ارائه نمی‌دهد و روابط رگرسیونی غیرخطی کاربرد بیشتری دارند. بنابراین با جایگزینی معادله رگرسیونی غیرخطی تعمیم‌یافته به جای رابطه (۲)، تابع کرنل $k(x, x_i)$ جایگزین بردار x می‌شود. معمولاً یافتن توابع کرنل خاصی که به‌عنوان بهترین عملگر برای آموزش مدل بردار پشتیبان استفاده شود، کاری دشوار است و لذا انتخاب توابع کرنل مناسب موضوع مهمی است که در یادگیری مدل‌های ماشین‌های بردار پشتیبان مطرح می‌شود. توابع کرنل ذکرشده در جدول ۱ کاربرد بیشتری جهت توسعه مدل‌های رگرسیونی بردار پشتیبان دارند (Smola and Scholkopf, 2004).

Table 1- Kernel functions of Support vector regression (Smola and Scholkopf, 2004)

جدول ۱- توابع کرنل مورداستفاده مدل‌های رگرسیونی بردار پشتیبان (Smola and Scholkopf, 2004)

Kernel function	Equation
Linear	$k(x, x_i) = x.x_i$
Polynomial	$k(x, x_i) = (1 + x.x_i)^n$
Gaussian (RBF)	$k(x, x_i) = exp(-\frac{\|x - x_i\|^2}{2\sigma^2}), \gamma = \frac{1}{2\sigma^2}$

۲-۲- به‌کارگیری راهکار فیلتر کالمن تجمیعی

داده‌گواری روشی است مبتنی بر توسعه و بهبود بخشیدن دقت مدل‌های قطعی که در علوم مختلف به‌طور فراوان استفاده شده است. به طور کلی، روش‌های داده‌گواری به بهنگام‌سازی و پردازش داده‌های پیش‌بینی‌شده توسط مدل‌های محاسباتی و شبیه‌سازی با توجه به داده‌های مشاهداتی جهت کاهش خطا و افزایش دقت مدل‌ها می‌پردازد. فیلتر کالمن از جمله راهکارهای فنون داده‌گواری بوده که هدف آن تخمین و پیش‌بینی مناسب مقدار نامعلوم x است، بدین ترتیب که با حداقل کردن واریانس مقادیر محتمل حول مقدار میانگین داده‌ها، مقادیر بهینه بدست می‌آیند. فیلتر کالمن به مرور زمان توسعه یافته و در قالب روش‌های تعمیم‌یافته فیلتر کالمن[۶] و روش فیلتر کالمن تجمیعی ارائه شده است. با علم به آنکه روابط ارائه شده جهت شبیه‌سازی فرایندهای سیستم‌های طبیعی از پیچیدگی بالایی برخوردار هستند، این روش‌ها توسعه یافته و در علوم هیدرولوژی به کار برده می‌شود (Drecourt, 2004).

Archive of SID.ir

Archive of SID.ir

روند بهنگام‌سازی داده‌ها با روش فیلتر کالمن تجمیعی بدین صورت است که فرض شود x یک بردار کنترل‌کننده برای یک نمونه شامل مجموعه‌ای از پارامترهای هیدرولیکی یا هیدرولوژیکی باشد. بردار x در گام زمانی t_{k+1} به عنوان بردار حالت پیشینه در زمان حال شناخته‌شده و به طور معمول با x^b نمایش داده می‌شود. حال چنانچه برای گام زمانی مذکور بردار مشاهدات و یا داده‌های واقعی (بردار y^0) موجود باشد که بتوان به‌وسیله آن بردار x^b را به‌روزرسانی کرد، لازم است در ابتدا بردار حالت x^b با بردار y هم جنس شود و یا به عبارتی به فضای بردار y انتقال داده شود که این کار توسط تابع عملگر $H(x^b)$ انجام می‌شود. سپس توسط توابع کرنل کالمن بردار x^b به بردار x^a بهنگام‌سازی می‌شود و در نهایت با استفاده از تابع عملگر $M(x^a)$، بردار حالت x^b برای گام زمانی t_{k+2} تولید می‌شود (Evensen, 2003). تابع کرنل بکار برده شده جهت بهنگام‌سازی داده‌های بردار حالت مطابق معادله (۴) تعریف می‌شود (Evensen, 2003):

$$x_i^a = x_i^b + K_i[y_i^o - H_i(x_i^b)] \quad (۴)$$

$$H_i = \frac{\partial y_i}{\partial x_i} = \frac{\partial H_i(x_i)}{\partial x_i} \quad (۵)$$

$$K_i = B_i H_i^T (H_i B_i H_i^T + R_i)^{-1} \quad (۶)$$

در روابط فوق اندیس i نشان‌دهنده گام‌های زمانی در به‌روزرسانی داده‌ها می‌باشد. ماتریس‌های B و R به ترتیب نمایانگر کواریانس خطاهای مربوط به داده‌های بردار حالت و مشاهداتی است. H ماتریس ژاکوبین عملگر $H(x^b)$ است که مقادیر عناصر آن از رابطه (۵) محاسبه می‌شود (Evensen, 2003). شایان‌ذکر است که پس از به‌روزرسانی داده‌ها در زمان i، عملگر $M_{i,i+1}$ داده‌ها را برای زمان $i+1$ پیش‌بینی می‌کند و این روند تا حداقل کردن واریانس خطاها پیش برده می‌شود.

$$x_{i+1}^b = M_{i,i+1}(x_i^a) \quad (۷)$$

۳- مطالعه موردی

حوضه آبریز قلعه‌شاهرخ-چلگرد در بالادست حوضه آبریز زاینده‌رود قرار دارد (شکل ۲) که بخش کوچکی از آن در استان اصفهان و مابقی در استان چهارمحال و بختیاری واقع شده است. مساحت زیر حوضه موردمطالعه ۱۴۹۸ کیلومترمربع است که سهم ارتفاعات ۱۲۸۹/۵ کیلومترمربع و دشت‌ها ۲۰۸/۵ کیلومترمربع است. شکل مذکور تغییرات رقوم ارتفاعی زیرحوضه مطالعاتی را نیز نشان می‌دهد.

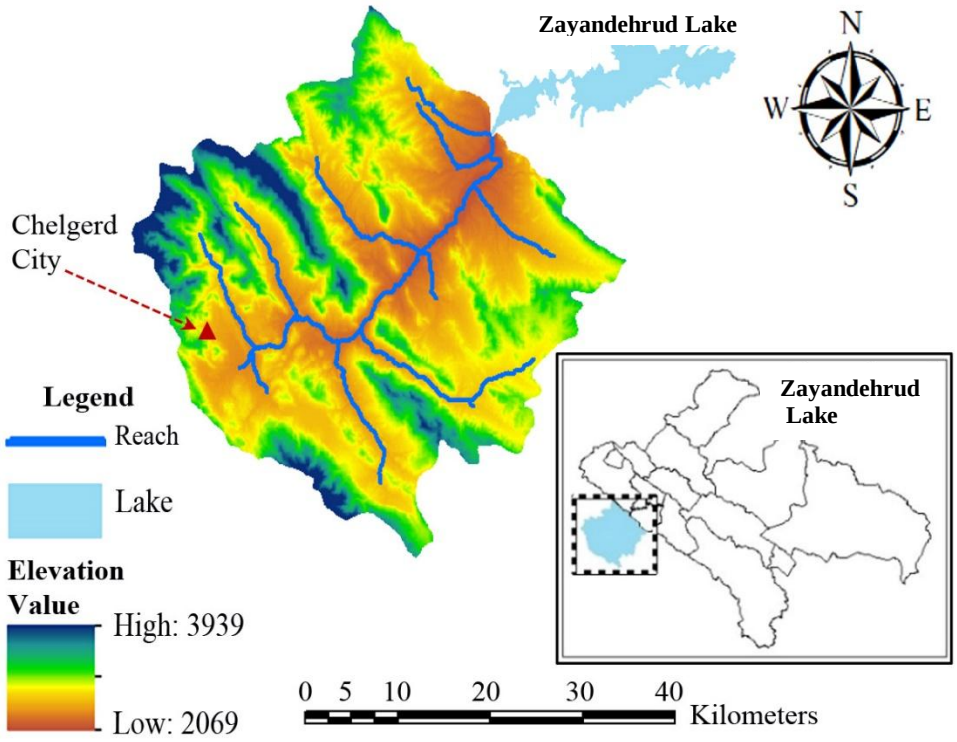


Fig. 2- Geographic location of Ghaleshahrokh- Chelgerd watershed

شکل ۲- نقشه موقعیت جغرافیایی حوضه آبریز قلعه‌شاهرخ-چلگرد

Archive of SID.ir

Archive of SID.ir

با استفاده از آمار و اطلاعات بارش سالانه ثبت‌شده در ایستگاه‌های همدید کوهرنگ و قلعه شاهرخ که به ترتیب به‌عنوان ایستگاه‌های اندازه‌گیری در نقاط مرتفع و پست حوضه مدنظر قرار می‌گیرند، میزان بارش سالانه به‌طور متوسط در ارتفاعات و دشت‌ها به ترتیب برابر ۱۲۸۲ و ۴۰۱ میلیمتر است. جدول ۲ تغییرات ماهانه متوسط بارش در محدوده حوضه مطالعاتی را نشان می‌دهد.

ایستگاه‌های آب‌سنجی در سطح حوضه‌های آبریز برای سنجش عمق آب، دبی جریان عبوری و اندازه‌گیری بار رسوبات عبوری در مقاطع مختلف آبراهه‌ها و یا رودخانه‌ها احداث می‌شوند. شکل ۳ موقعیت ایستگاه‌های آب‌سنجی محدوده مطالعاتی را نشان می‌دهد.

Table 2- Monthly long-time precipitation of Chelgerd sub-basin (mm)
جدول ۲- میانگین درازمدت مجموع بارش ماهانه منطقه مطالعاتی قلعه‌شاهرخ-چلگرد (میلیمتر)

Month	October	November	December	January	February	March	April	May	June	July	August	September	Annual
Height	41	177	202	214	202	239	159	43	1	1	1	2	1282
plain	13	53	67	66	57	70	52	17	3	1	1	1	401



Fig. 3- Location of hydrometric gauges in Ghaleshahrokh- Chelgerd watershed
شکل ۳- موقعیت ایستگاه‌های آب‌سنجی حوضه آبریز قلعه‌شاهرخ-چلگرد

Archive of SID.ir

Archive of SID.ir

در زیر حوضه آبریز مطالعاتی انتقال آب بین حوضه‌ای توسط دو تونل اول و دوم کوهرنگ در بالادست حوضه انجام می‌شود که ایستگاه‌های چلگرد و دره‌در به ترتیب به منظور اندازه‌گیری دبی جریان ورودی به حوضه از تونل اول و دوم احداث شدند. آبدهی تونل اول کوهرنگ به داخل حوضه از سال ۱۳۲۶ شمسی آغاز شده است. این تونل در حومه شهرستان چلگرد واقع شده است. به دلیل مشکلات فنی و انسانی، ایستگاه دبی سنج تونل اول از مهرماه ۱۳۷۵ (اکتبر ۱۹۹۶) تا شهریورماه ۱۳۷۸ (دسامبر ۱۹۹۹) دارای خلاء آماری است. سری زمانی داده‌های ثبت‌شده ایستگاه چلگرد مطابق نمودار مندرج در شکل (۴-الف) است. آبدهی تونل دوم کوهرنگ به داخل حوضه از سال ۱۳۶۶ شمسی آغاز شده است. سری زمانی دبی آبدهی تونل دوم برای دوره مطالعاتی مطابق شکل (۴-ب) است که از مهرماه ۱۳۸۳ (اکتبر ۲۰۰۴) دارای خلاء آماری است. ایستگاه اندازه‌گیری تونل دوم به دلیل ساخت نیروگاه در پایین‌دست تونل و انحراف مسیر آب به سمت نیروگاه تعطیل شده و بعد از بازه زمانی مذکور، ایستگاه دره نمکی در شهریورماه ۱۳۸۷ (دسامبر ۲۰۰۸) راه‌اندازی شده که با کسر آبدهی تونل اول از آبدهی ایستگاه دره نمکی، مقدار آبدهی تونل دوم تخمین زده شد.

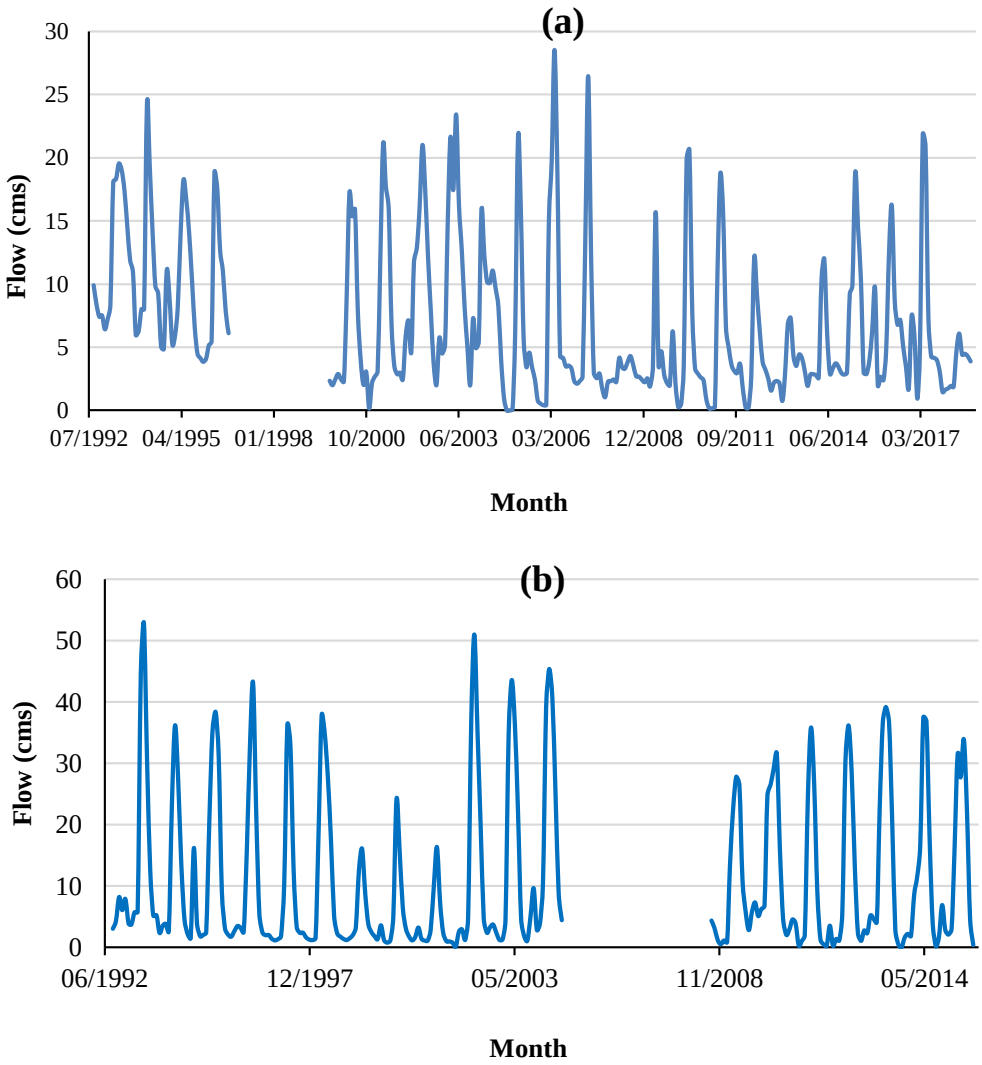


Fig. 4- (a) Discharge time series of Chelgerd gauge (1st tunnel) for 1992-2018, (b) Discharge time series of Dare Dor and Dare Namaki gauges (2nd tunnel) for 1992-2018

شکل ۴- (الف) سری زمانی آبدهی ایستگاه آب‌سنجی چلگرد (تونل اول) برای دوره آماری ۹۷-۱۳۷۱، (ب) سری زمانی آبدهی ایستگاه‌های آب‌سنجی دره در و دره نمکی (تونل دوم) برای دوره آماری ۹۴-۱۳۷۱

Archive of SID.ir

Archive of SID.ir

۴- تحلیل نتایج

با توجه به اطلاعات ارائه شده در بخش قبلی، خلاء آماری داده‌های دبی سنجی در سطح حوضه مطالعاتی به دلیل تازه تأسیس بودن ایستگاه‌های نامبرده، تعطیلی ایستگاه و یا خرابی ایستگاه مشهود است. به‌خصوص، چنانچه هدف بررسی آورد تونل‌های انتقال آب حوضه درگذشته باشد، خلاء آماری در ایستگاه‌های مذکور موجب ناتوانی در بررسی و تحلیل آماری می‌شود. از این رو، در این مطالعه با انتخاب ایستگاه چلگرد به‌عنوان مطالعه موردی، به بازسازی داده‌های ایستگاه نامبرده پرداخته شد. لازم به ذکر است بازه آماری موردمطالعه با توجه به آخرین آمار کسب‌شده از سازمان‌های آب منطقه‌ای و هواشناسی، از سال آبی ۱۳۷۱ (۱۹۹۲ میلادی) الی ۱۳۹۷ (۲۰۱۸ میلادی) انتخاب شده است. شکل ۵، شمایی از موقعیت ایستگاه‌های موردمطالعه در حوضه آبریز مطالعاتی قبل از سال ۱۳۸۳ (۲۰۰۴ میلادی) و بعد از سال ۱۳۸۷ (۲۰۰۸ میلادی) را نشان می‌دهد. همان‌طور که مشاهده می‌شود در ایستگاه دره در، خروجی دبی آورد تونل دوم کوهرنگ اندازه‌گیری می‌شود، ولی در ایستگاه دره نمکی خروجی مجموع دو تونل اول و دوم اندازه‌گیری می‌شود. با توجه به این مسئله که از سال ۱۳۸۷، ایستگاه دره نمکی احداث‌شده، دبی تونل دوم با کسر مقادیر دبی اندازه‌گیری ایستگاه چلگرد از ایستگاه دره نمکی قابل محاسبه است. لازم به ذکر است به دلیل نزدیکی موقعیت ایستگاه‌های موردنظر، دبی ورودی و خروجی در طول مسیر ناشی از عوامل دیگر نظیر بارش، چشمه، برداشت (جهت مصارف) قابل صرف‌نظر است. بنابراین، فرض بیان‌شده مبنی بر محاسبه دبی آورد تونل دوم از سال ۱۳۸۷ به بعد (بعد از ۲۰۰۸ میلادی) قابل توجیه است.

به منظور تولید داده‌های دبی ایستگاه چلگرد که دارای خلاء آماری است، از مدل رگرسیونی ماشین‌های بردار پشتیبان (SVR) استفاده شد که بتوان به‌طور تقریبی داده‌هایی که به دلیل مشکلات فنی اندازه‌گیری نشده است را تولید کرد. چنانچه ورودی مدل داده محور، دبی آبدهی تونل در دو ماه گذشته (Q_{t-1}, Q_{t-2}) و بارش در ماه حاضر باشد و خروجی مدل دبی آبدهی در ماه حاضر (Q_t) باشد، می‌توان یک مدل داده محور با توجه به داده‌های دبی موجود تونل اول و بارش اندازه‌گیری شده ایستگاه همدید کوهرنگ آموزش داد و آن مدل را آزمایش کرد. در محیط MATLAB و با استفاده از کدهای دستوری مدل SVR، مدل‌سازی انجام شد. در مجموع ۲۷۸ داده تاریخی در دسترس بوده که از این آمار ۲۱۰ ماه (۷۶ درصد) داده آماری برای مرحله آموزش و ۶۸ ماه داده آماری برای مرحله آزمون (۲۴ درصد) در نظر گرفته شده است. مدل‌سازی با استفاده از توابع کرنل RBF انجام شده، که علاوه بر پارامترهای مدل SVR نظیر ضریب جریمه C و میزان انحراف ε، پارامتر γ نیز به‌عنوان پارامترهای قابل تنظیم جهت بهبود عملکرد مدل در مرحله آموزش قابل تغییر است.

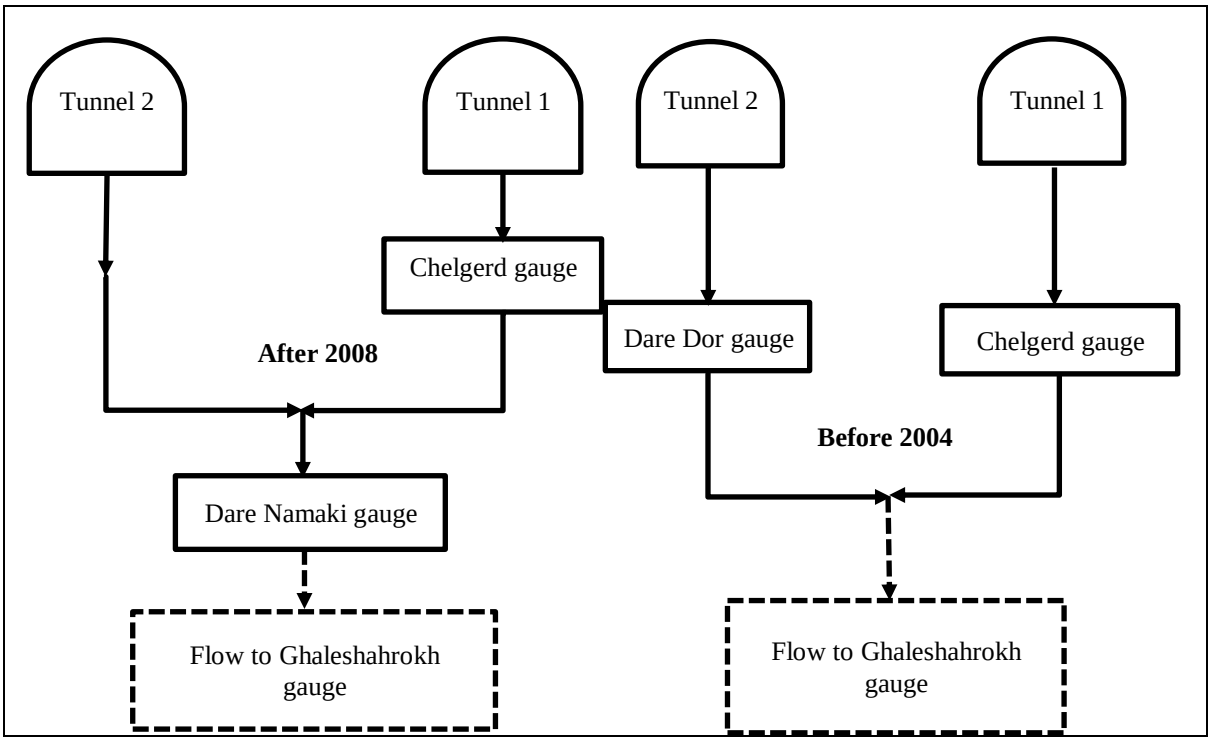


Fig. 5- Scheme of hydrometric gauges before 2004 and after 2008

شکل ۵- شمایی از موقعیت ایستگاه‌های آب‌سنجی مطالعاتی قبل از سال ۱۳۸۳ و بعد از سال ۱۳۸۷ شمسی

Archive of SID.ir

Archive of SID.ir

بنابراین واسنجی مدل SVR در این مطالعه با تنظیم سه پارامتر ذکرشده انجام شد. انتخاب مقادیر بهینه پارامترهای مذکور با روش آزمون و خطا صورت گرفت و بهترین نتایج در مرحله آموزش با ضریب همبستگی ۰/۸۳ بدست آمد. نمودار مندرج در شکل ۶ مقایسه داده‌های شبیه‌سازی و مشاهداتی را در مرحله آزمون با فرض مقادیر $C = 10$، $\varepsilon = 0.1$ و $\gamma = 10$ نشان می‌دهد.

پس از مرحله آموزش، به‌منظور صحت‌سنجی مدل ساخته‌شده، مقادیر پارامترهای C، ε و γ تثبیت شده و مدل مورد نظر برای داده‌هایی که جهت مرحله آزمون ذخیره شده، اجرا شد. نتایج بدست آمده از مرحله آزمون در مقایسه با داده‌های اندازه‌گیری شده توسط ایستگاه آب‌سنجی در شکل ۷ نشان داده شده است. مطابق نتایج بدست آمده، ضریب همبستگی در مرحله آزمون برابر ۰/۷ حاصل شده است.

مدل رگرسیونی با ضریب $R = 0.83$ (در مرحله آموزش) و $R = 0.7$ (در مرحله آزمون) جهت تخمین دبی ماهانه ایستگاه استفاده می‌شود. شکل ۸ نتایج تخمین و یا تولید داده‌های ثبت نشده ایستگاه چلگرد در بازه زمانی مهرماه ۱۳۷۵ (اکتبر ۱۹۹۶) تا شهریورماه ۱۳۷۸ (دسامبر ۱۹۹۹) را نشان می‌دهد. همان‌طور که مشاهده می‌شود، روند سری زمانی شبیه‌سازی‌شده در بازه زمانی مذکور با روند تاریخی داده‌های اندازه‌گیری شده در طول دوره زمانی همخوانی ندارد. بنابراین، لازم است عملکرد مدل رگرسیونی بهبود و اصلاح یابد. بدین منظور از راهکار فیلتر کالمن تجمیعی جهت اصلاح مدل رگرسیونی استفاده می‌شود.

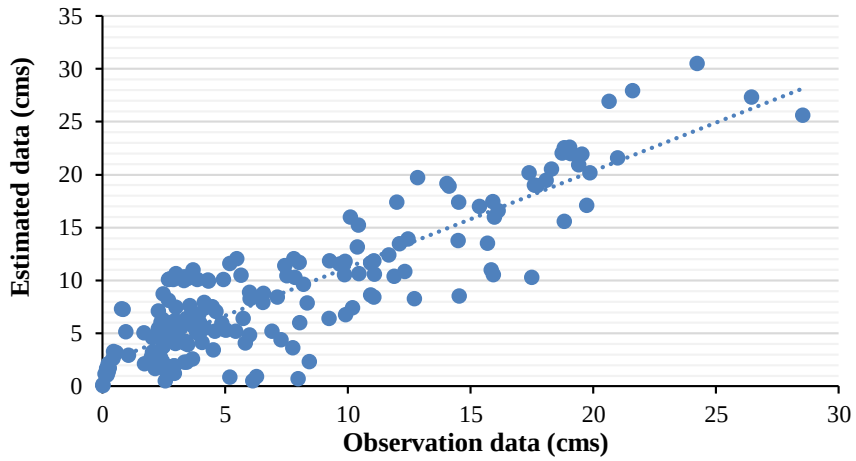


Fig. 6- Comparison of SVR estimated and observation discharge data of Chelgerd gauge (Training stage)

شکل ۶- مقایسه داده‌های تخمینی مدل SVR و مشاهداتی دبی ایستگاه آب‌سنجی چلگرد (در مرحله آموزش)

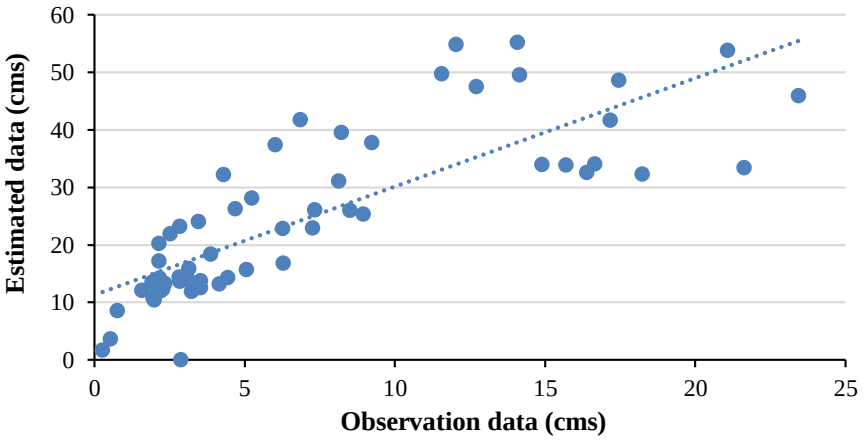


Fig. 7- Comparison of SVR estimated and observation discharge data of Chelgerd gauge (Testing stage)

شکل ۷- مقایسه داده‌های تخمینی مدل SVR و مشاهداتی دبی ایستگاه آب‌سنجی چلگرد (در مرحله آزمایش)

Archive of SID.ir

Archive of SID.ir

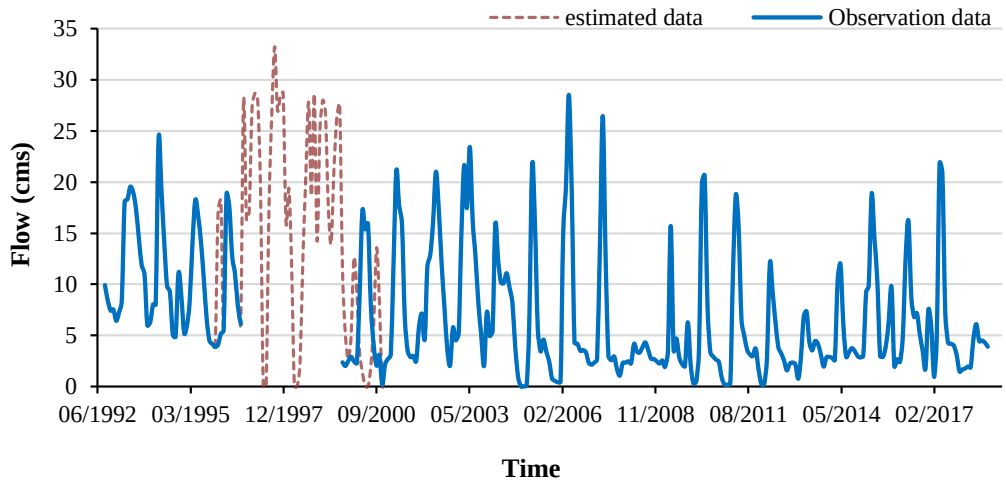


Fig. 8- Comparison of time series of Chelgerd estimated and observation discharge data
شکل ۸- مقایسه سری زمانی داده‌های دبی تخمینی و مشاهداتی چلگرد

به منظور اصلاح داده‌های تولیدشده از راهکار فیلتر کالمن تجمیعی، بایستی از یک ایستگاه شاهد جهت کاهش خطا استفاده کرد. ایستگاه آب‌سنجی قلعه‌شاهرخ که در خروجی زیر حوضه واقع است به‌عنوان ایستگاه شاهد انتخاب شد. از آنجا که موقعیت مکانی تونل اول و ایستگاه قلعه شاهرخ با یکدیگر همخوانی ندارند، یک مدل داده محور دوم برای برقراری ارتباط بین این دو ایستگاه توسعه یافت. بدین منظور ایستگاه آب‌سنجی دره نمکی واقع در پایین‌دست دو تونل اول و دوم انتخاب شد که داده‌های اندازه‌گیری توسط این ایستگاه مجموع دبی‌های تونل اول و دوم کوهرنگ است. به همین ترتیب ورودی مدل داده محور دوم دبی ایستگاه دره نمکی و خروجی آن دبی ایستگاه قلعه‌شاهرخ است. سری زمانی دبی اندازه‌گیری شده در ایستگاه دره نمکی و ایستگاه قلعه‌شاهرخ مطابق شکل ۹ است. شایان‌ذکر است، خلاء آماری در سری زمانی داده‌های دره نمکی به دلیل عدم اندازه‌گیری داده‌های دبی ایستگاه چلگرد (در بازه زمانی اکتبر سال ۱۹۹۶ الی آگوست ۱۹۹۹) و عدم اندازه‌گیری دبی ایستگاه دره‌در (در بازه زمانی اکتبر ۲۰۰۴ الی اکتبر ۲۰۰۸) است.

عملکرد دو مدل رگرسیونی توسعه یافته با استفاده از شاخص‌های ارزیابی خطا نظیر ضریب همبستگی (R)، ضریب نش–ساتکلیف (NS^7) و معیار ریشه متوسط مربعات خطا (RMSE) سنجیده شدند که جدول ۳ نتایج حاصل از ارزیابی را نشان می‌دهد.

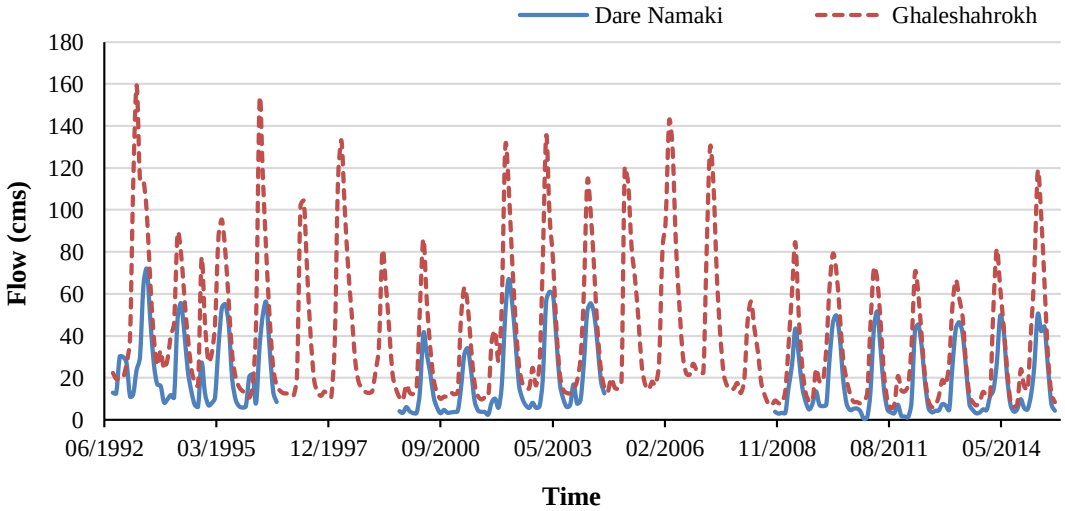


Fig. 9- Time series of Dare Namaki and Ghaleshahrokh observation discharge data
شکل ۹- سری زمانی دبی عبوری مشاهداتی از ایستگاه‌های آب‌سنجی قلعه‌شاهرخ و دره نمکی

Archive of SID.ir

Archive of SID.ir

ساختار روش پیشنهادی جهت اصلاح داده‌های شبیه‌سازی با راهکار فیلتر کالمن تجمیعی در شکل ۱۰ نشان داده شده است دبی جریان در ایستگاه چلگرد (خروجی تونل اول) با استفاده از راهکار EnKF بازتولید و اصلاح شد. با توجه به الگوریتم مذکور، دبی ایستگاه چلگرد به‌عنوان بردار حالت سیستم (X^b)، دبی مشاهداتی ایستگاه قلعه‌شاهرخ به‌عنوان بردار مشاهدات (y) و دبی شبیه‌سازی‌شده قلعه‌شاهرخ به‌عنوان بردار تبدیل یافته بردار حالت به جنس بردار مشاهدات ($H(X^b)$) شناخته می‌شود.

Table 3- Checking results of SVR models by evaluation error criteria

جدول ۳- نتایج ارزیابی مدل‌های SVR ساخته‌شده با شاخص‌های ارزیابی خطا

Train step			
NS	RMSE	R	SVR models
0.71	3.42	0.83	Chelgerd model
0.81	5.20	0.96	Ghaleshahrokh model
Test step			
NS	RMSE	R	SVR models
0.25	20.38	0.70	Chelgerd model
0.66	6.60	0.76	Ghaleshahrokh model

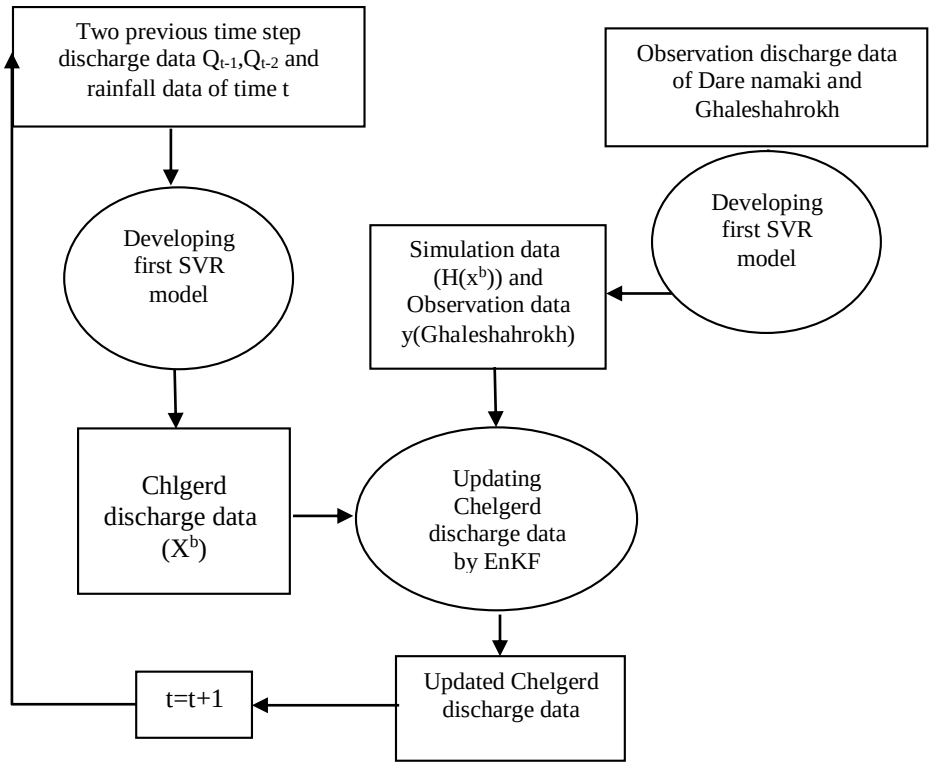


Fig. 10- Data assimilation algorithm for updating and improving estimated discharge data of Chelgerd gauge

شکل ۱۰- الگوریتم راهکار داده‌گواری جهت اصلاح دبی تخمینی ایستگاه چلگرد

پس از بازتولید و اصلاح مقادیر شبیه‌سازی‌شده ایستگاه چلگرد، سری زمانی آن مطابق با نمودار شکل ۱۱ است. به‌منظور مقایسه عملکرد ابزار مورداستفاده، در نمودار شکل ۱۲، سری زمانی داده‌های تولیدشده برای بازه زمانی بدون داده، بر اساس نتایج مدل SVR با/بدون استفاده از راهکار EnKF، به‌صورت مقایسه‌ای نشان داده شده است. عملکرد راهکار EnKF به صورتی است که در هر مرحله شبیه‌سازی به صورت پیوسته و گام‌به‌گام واریانس خطاهای داده‌های شبیه‌سازی کاهش و در نتیجه انحراف از میانگین داده‌های جامعه آماری تقلیل می‌یابد. به‌عنوان مثال در نمودار شکل مذکور، مقدار دبی شبیه‌سازی‌شده (غیر اصلاحی) در ماه آگوست ۱۹۹۷ برابر ۲۶/۸ مترمکعب بر ثانیه است و پس از بازسازی (با استفاده از راهکار EnKF)، دبی به مقدار ۱/۲ مترمکعب بر ثانیه اصلاح یافت. با توجه به آن که ماه آگوست تقریباً نشان‌دهنده نیمه دوم مرداد ماه است و میزان بارش در این ماه به حداقل مقدار خود می‌رسد، بنابراین دبی ۱/۲ مترمکعب بر ثانیه نسبت به ۲۶/۸ مترمکعب بر ثانیه به واقعیت نزدیک‌تر است. این اختلاف مقدار، نشان‌دهنده انحراف زیاد نتایج شبیه‌سازی اولیه (غیراصلاحی) از میانگین داده‌های جامعه آماری برای گام زمانی مورد نظر است که به

Archive of SID.ir

Archive of SID.ir

موجب اصلاح داده‌ها با راهکار مورد نظر، در برخی موارد منجر به تغییر روند و یا رفتار سری زمانی می‌شود.

با وجود اینکه داده‌های اندازه‌گیری ایستگاه چلگرد برای سال آماری (سپتامبر سال ۹۹ الی اکتبر ۲۰۰۲) موجود است، شبیه‌سازی با مدل SVR برای بازه زمانی مذکور به منظور صحت‌سنجی ادامه و انجام یافته است. همانطور که مشاهده شد، علیرغم عملکرد مناسب مدل SVR (با سنجش شاخص‌های ارزیابی خطا)، مدل SVR به تنهایی برای شبیه‌سازی و تولید داده عملکرد مناسبی در تولید داده‌های بازه زمانی مذکور در مقایسه با داده‌های اندازه‌گیری از خود نشان نداده است. به همین دلیل اهمیت به‌کارگیری راهکار فیلتر کالمن تجمیعی (EnKF) در این زمینه قابل توجه است. نمودار مندرج در شکل ۱۳ اصلاح داده‌های شبیه‌سازی شده مدل SVR برای بازه زمانی مذکور را نشان می‌دهد. همانطور که مشاهده می‌شود پس از اصلاح داده‌ها (با راهکار EnKF)، داده‌های اصلاح یافته همپوشانی مناسبی با داده‌های اندازه‌گیری (مشاهداتی) از خود نشان داده‌اند.

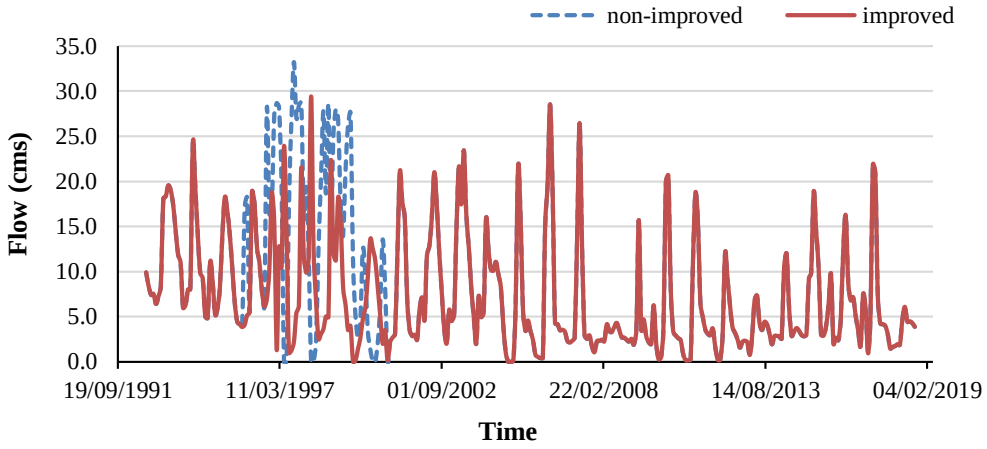


Fig. 11- Time series of improved and non-improved discharge data of Chelgerd gauge

شکل ۱۱- سری زمانی اصلاح یافته و نیافته آبدهی ایستگاه چلگرد

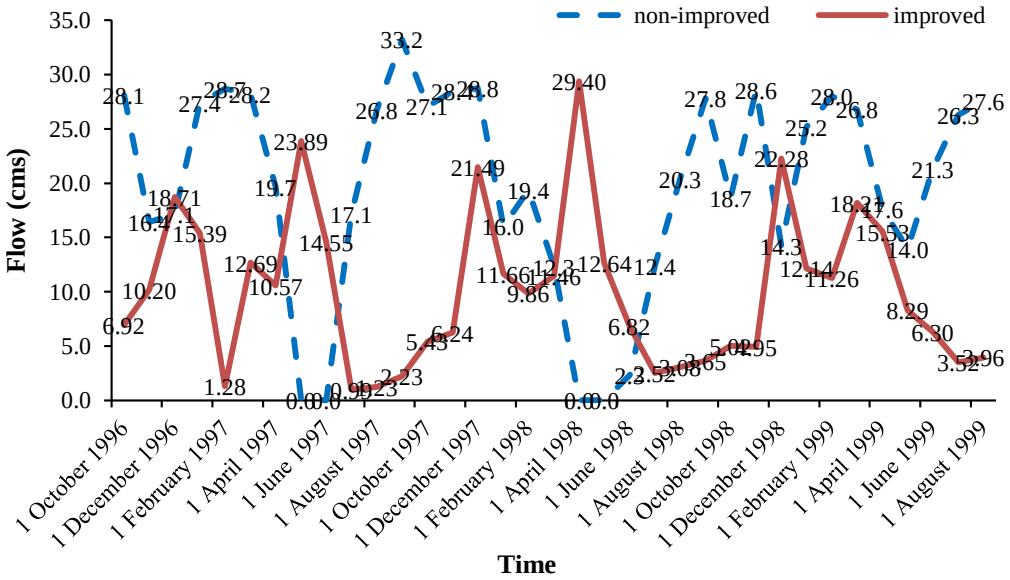


Fig. 12- Comparison of time series of improved and non-improved estimated discharge data of Chelgerd for the no-obs-data period

شکل ۱۲- مقایسه سری زمانی اصلاحی و غیر اصلاحی آبدهی ایستگاه چلگرد برای بازه زمانی بدون داده

Archive of SID.ir

Archive of SID.ir

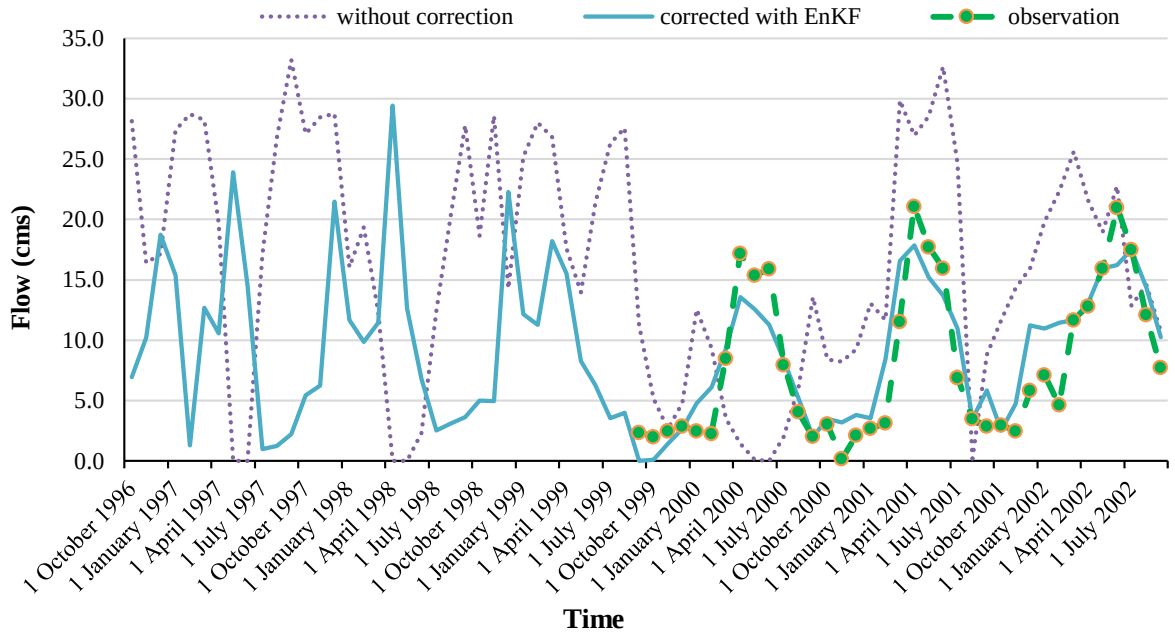


Fig. 13-Validation and comparison of improved model with non-improved model for generation discharge data of Chelgerd

شکل ۱۳- صحت‌سنجی مدل اصلاحی در مقایسه با مدل غیر اصلاحی به منظور تولید داده‌های ایستگاه چلگرد

به منظور ارزیابی کمّی نتایج، از شاخص‌های ارزیابی خطا استفاده شده که نتایج آن‌ها برای دو مدل غیراصلاحی و اصلاحی در جدول ۴ ارائه شده است. نتایج بدست آمده از ارزیابی شاخص‌های خطا در جدول مذکور با توجه به موجود بودن داده‌های مشاهداتی برای سال‌های ۲۰۰۰ الی ۲۰۰۲ (بازه ۳۶ ماهه در بازه زمانی صحت‌سنجی)، ارائه شده است. همانطور که مشاهده می‌شود مقادیر R=۰/۴ و NS=-۰/۲۶ نشان دهنده عملکرد نامناسب مدل رگرسیونی غیر اصلاحی (قبل از بکار گیری راهکار EnkF) است. همانطور که پیش‌تر نیز اشاره شده، ضریب همبستگی R نشان‌دهنده کارآمدی مدل و بطور کلی بررسی روند تغییرات نسبت داده‌های مشاهداتی است و بر خلاف روش NS به داده‌های اوج و حداقل در یک سری زمانی حساس نیست. با این وجود زمانیکه هر دو این ضرایب مقادیر کمی را از خود نشان دهند، بطور کلی عملکرد مدل مناسب ارزیابی نمی‌شود که این مورد مصداق نتایج بدست آمده برای مدل غیر اصلاحی است. دلیل این مسأله وجود چالش‌های متعدد از جمله برداشت‌های به حساب نیامده از منابع آب سطحی و زیرزمینی است. به عنوان مثال برداشت غیر مجاز از چاه‌ها، برداشت‌های غیر مجاز از منابع آب سطحی و عدم دسترسی به آمار مصارف آب توسط صنایع سبب شده این حوضه آبریز از حالت طبیعی خود خارج شود. بنابراین در این مطالعه اصلاح عملکرد مدل‌ها به کمک راهکار فیلتر کالمن تجمیعی تحقق یافت. با استفاده از این راهکار، در هر گام زمانی از شبیه‌سازی جهت تخمین داده ثبت نشده، نتایج شبیه‌سازی بهبود یافته و داده‌های به‌روزرسانی شده برای شبیه‌سازی گام زمانی بعدی مورد استفاده قرارگرفت و این امر در طی شبیه‌سازی‌های متوالی، سبب بهبود عملکرد مدل‌سازی شده است.

چنانچه به نتایج مدل اصلاحی توجه شود، مقادیر R=۰/۸۹ و NS=۰/۷۰ نشان‌دهنده کارآمدی مدل و نتایج مناسب بدست آمده برای نقاط اوج و حداقل سری زمانی داده‌ها است. علاوه بر شاخص‌های مذکور، کم‌تر بودن میزان خطای شبیه‌سازی (با ارزیابی شاخص RMSE) برتری مدل اصلاحی را نسبت به غیر اصلاحی نشان می‌دهد.

Table 4- Checking results of SVR models without and with EnKF by evaluation error criteria for validation stage

جدول ۴- نتایج ارزیابی مدل‌های غیر اصلاحی و اصلاحی چلگرد با شاخص‌های ارزیابی خطا در مرحله صحت‌سنجی

NS	RMSE	R	SVR models
-0.26	10.06	0.4	Modelling without EnKF
0.7	2.94	0.89	Modelling with EnKF

Archive of SID.ir

Archive of SID.ir

۵- جمع‌بندی

در این مطالعه با هدف تولید داده‌های ثبت‌نشده دبی جریان آب در یک نقطه از حوضه آبریز قلعه‌شاهرخ –چلگرد، از مدل رگرسیونی بردار پشتیبان (SVR) استفاده شد. با جمع‌آوری داده‌های موردنیاز شامل داده‌های بارش و داده‌های اندازه‌گیری ایستگاه‌های آب‌سنجی در نقاط موردنظر، مدل داده محور در دو مرحله آموزش و آزمون ساخته و صحت‌سنجی شد. مدل رگرسیونی SVR، جهت تولید داده‌های ثبت‌نشده ایستگاه چلگرد که در محل خروجی تونل اول کوهرنگ احداث شده است، ساخته و توسعه یافت. مدل موردنظر با ضریب همبستگی ۰/۸۳ و ۰/۷ به ترتیب در مرحله آموزش و آزمون عملکرد قابل‌قبولی از خود نشان داد. عملکرد مدل رگرسیونی ساخته‌شده با استفاده از شاخص‌های ارزیابی RMSE و نش–ساتکلیف (NS) نیز سنجیده شد که برای مرحله آموزش، به ترتیب مقادیر ۳/۴۲ مترمکعب بر ثانیه و ۰/۷۱ بدست آمد. مقادیر مذکور برای مرحله آزمون به ترتیب برابر ۲۰/۳۸ مترمکعب بر ثانیه و ۰/۲۵ بدست آمد. به‌منظور بهبود عملکرد مدل موردنظر، از راهکار فیلتر کالمن تجمیعی (EnKF) از دسته فنون داده‌گواری، استفاده شد. راهکار یادشده با به‌روزرسانی داده‌های تخمینی در هر گام زمانی، به‌طور متوالی عملکرد مدل رگرسیونی را در طول بازه زمانی بهبود داده است. عملکرد راهکار داده‌گواری مورداستفاده بدین صورت است که به کمک داده‌های ایستگاه شاهد با جامعه آماری قابل‌قبول، واریانس انحراف داده‌های تخمینی و یا محاسباتی را از میانگین جامعه آماری کاهش می‌دهد. بدین منظور از ایستگاه آب‌سنجی قلعه‌شاهرخ واقع در خروجی حوضه آبریز استفاده شد. برای یک بعد کردن و ایجاد یک ارتباط معنایی بین داده ایستگاه ثبت نشده و ایستگاه شاهد، از یک مدل رگرسیونی SVR دومی استفاده شد. مدل موردنظر دوم نیز با ضریب همبستگی ۰/۹۶ و ۰/۷۶ به ترتیب برای مرحله آموزش و آزمون توسعه یافت. مقادیر شاخص ارزیابی RMSE و NS برای مرحله آموزش مدل دوم به ترتیب برابر ۵/۲ مترمکعب بر ثانیه و ۰/۸۱ محاسبه شده است و برای مرحله آزمون، این مقادیر به ترتیب برابر ۶/۶ مترمکعب بر ثانیه و ۰/۶۶ بدست آمده است.

شبیه‌سازی با مدل SVR برای بازه زمانی ۳۶ ماهه سال ۲۰۰۰ تا ۲۰۰۲ به منظور صحت‌سنجی ادامه و انجام یافت. نتایج برای مدل رگرسیونی تلفیقی با راهکار فیلتر کالمن تجمیعی نشان داد که پس از اصلاح داده‌ها (با بکارگیری راهکار EnKF)، داده‌های اصلاح یافته همپوشانی مناسبی با داده‌های اندازه‌گیری (مشاهداتی) از خود نشان داده‌اند که با ارزیابی شاخص‌های ارزیابی خطا، ضریب همبستگی برابر R=۰/۸۹، ضریب نش برابر NS=۰/۷۰ و شاخص RMSE برابر ۲/۹۴ مترمکعب بر ثانیه بدست آمد.

پی‌نوشت‌ها

1- Data Assimilation Technics
2- Support Vector Machine
3- Root Mean Square Error
4- Support Vector Regression
5- Ensemble Kalman Filter (EnKF)
6- Extended Kalman Filter (EKF)
7- Nash-Sutcliffe

Archive of SID.ir

Archive of SID.ir

۶- مراجع

Asgharinia S, Petroselli A (2020) A comparison of statistical methods for evaluating missing data of monitoring wells in the Kazeroun Plain, Fars Province, Iran. Groundwater for Sustainable Development 10:100294

Andreadis K M, Lettenmaier D P (2006) Assimilating remotely sensed snow observations into a macroscale hydrology model. Advances in Water Resources 29(6):872-886

Bosilovich M G, Radakovich J D, da SILVA A, Todling R, Verter F (2007) Skin temperature analysis and bias correction in a coupled land-atmosphere data assimilation system. Journal of the Meteorological Society of Japan Ser. II 85:205-228

Coulibaly P, Baldwin C K (2005) Nonstationary hydrological time series forecasting using nonlinear dynamic methods. Journal of Hydrology 307(1-4):164-174

Crow W T, Wood E F (2003) The assimilation of remotely sensed soil brightness temperature imagery into a land surface model using ensemble Kalman filtering: A case study based on ESTAR measurements during SGP97. Advances in Water Resources 26(2):137-149

Dastorani M T, Moghadamnia A, Piri J, Rico-Ramirez M (2010) Application of ANN and ANFIS models for reconstructing missing flow data. Environmental monitoring and assessment 166(1):421-434

Dikbas F, Yasar M (2020) Data-driven modeling of flows of Antalya basin and reconstruction of missing data. Iranian Journal of Science and Technology, Transactions of Civil Engineering 44(4):1335-1344

Hirsch R M (1979) An evaluation of some record reconstruction techniques. Water Resources Research 15(6):1781-1790

Jadidoleslam N, Mantilla R, Krajewski W F (2021) Data assimilation of satellite-based soil moisture into a distributed hydrological model for streamflow predictions. Hydrology 8(1):52

Kashif Gill M, Kemblowski M W, McKee M (2007) Soil moisture data assimilation using support vector machines and ensemble Kalman filter1. Journal of the American Water Resources Association 43(4):1004-1015

Khalil M, Panu U S, Lennox W C (2001) Groups and neural networks based streamflow data infilling procedures. Journal of Hydrology 241(3-4):153-176

Kuligowski R J, Barros A P (1998) Using artificial neural networks to estimate missing rainfall data 1. Journal of the American Water Resources Association 34(6):1437-1447

Kumar S V, Reichle R H, Peters-Lidard C D, Koster R D, Zhan X, Crow W T, Houser P R (2008) A land surface data assimilation framework using the land information system: Description and applications. Advances in Water Resources 31(11):1419-1432

Langhammer J, Česák J (2016) Applicability of a nu-support vector regression model for the completion of missing data in hydrological time series. Water 8(12):560

Li X L, Lü H, Horton R, An T, Yu Z (2014) Real-time flood forecast using the coupling support vector machine and data assimilation method. Journal of Hydroinformatics 16(5):973-988

Liu D, Yu Z B, Lue H S (2010) Data assimilation using support vector machines and ensemble Kalman filter for multi-layer soil moisture prediction. Water Science and Engineering 3(4):361-377

McCulloch J A W, Booth M (1970) Estimation of basin precipitation by regression equation. Water Resources Research 6(6):1753-1758

Mehrparvar M, Asghari K (2018) Modular optimized data assimilation and support vector machine for hydrologic modeling. Journal of Hydroinformatics 20(3):728-738

Mehrparvar M, Asghari K, Golmohammadi M (2019) Reducing error of rainfall-runoff simulation using coupled hydrological SWAT Model and data assimilation technique. Iran-Water Resources Research 14(5):85-102 (In Persian)

Noori R, Karbassi A R, Moghaddamnia A, Han D, Zokaei-Ashtiani M. H, Farokhnia A, Gousheh M G (2011) Assessment of input variables determination on the SVM model performance using PCA, Gamma test, and forward selection techniques for monthly stream flow prediction. Journal of hydrology 401(3-4):177-189

Petty T R, Dhingra P (2018) Streamflow hydrology estimate using machine learning (SHEM). Journal of the American Water Resources Association 54(1):55-68

Raman H, Mohan S, Padalinathan P (1995) Models for extending streamflow data: A case study. Hydrological Sciences Journal 40(3):381-393

Sattari M T, Falsafian K, Irvem A, Qasem S N (2020) Potential of kernel and tree-based machine-learning models for estimating missing data of rainfall. Engineering Applications of Computational Fluid Mechanics 14(1):1078-1094

Archive of SID.ir

Archive of SID.ir

Sattari M T, Rezazadeh-Joudi A, Kusiak A (2017) Assessment of different methods for estimation of missing data in precipitation studies. Hydrology Research 48(4):1032-104

Smola A J, Schölkopf B (2004) A tutorial on support vector regression. Statistics and Computing 14(3):199-222

Vapnik V (1995) The nature of statistical learning theory. New York, Springer

Wallis J R, Lettenmaier D P, Wood E F (1991) A daily hydroclimatological data set for the continental United States. Water Resources Research 27(7):1657-1663

Zhu Q, Wang Y, Luo Y (2021) Improvement of multi-layer soil moisture prediction using support vector machines and ensemble Kalman filter coupled with remote sensing soil moisture datasets over an agriculture dominant basin in China. Hydrological Processes 35(4):e14154

Archive of SID.ir