

Evaluation of Moving Average Pre-processing Approach to Improve the Efficiency of Support Vector Regression Model for Inflow Prediction

MAHDI ABBASI¹, SHAHAB ARAGHINEJAD^{2*}, KUMARS EBRAHIMI³

1. MSc in Water Resources Engineering, Department of Irrigation & Reclamation Engineering, University of Tehran, Karaj, Iran
 2. Associate Professor, Department of Irrigation and Reclamation Engineering, University of Tehran, Karaj, Iran
 3. Professor, Department of Irrigation and Reclamation Engineering, University of Tehran, Karaj, Iran.
- (Received: March. 27, 2017- Revised: June. 23, 2018- Accepted: July. 3, 2018)

ABSTRACT

Accurate hydrological forecasting is a main tool for the water resources planning. In this paper, the inflow rates to Bakhtiari and Rudbar Dams in Lorestan province – IRAN, were forecasted using support vector regression (SVR), Multiple Linear Regression (MLR) and Autoregressive Moving Average (ARMA) models. In order to pre-process the input data for the above mentioned models, the moving average approach was used. Furthermore, Nash-Sutcliffe Efficiency (NSE), Root Mean Square Error (RMSE), correlation coefficient (R) and Taylor diagram were used to evaluate the efficiency of the models. The results showed that the moving average pre-processing approach improved the performance of the above mentioned models dramatically. For instance, the values of Nash-Sutcliffe correspond to SVR hybrid model in forecasting inflow rate to Bakhtiari and Rudbar-Lorestan dams with moving average pre-processing were improved by 13.4% and 6.6%, respectively, as compared to those in the SVR model without pre-processing.

Keywords: Forecasting Time Series, Moving Average, Bakhtiari Dam, Rudbar Dam, ARMA

* Corresponding Author's Email: Araghinejad@ut.ac.ir

ارزیابی رویکرد پیش‌پردازش میانگین متحرک در تدقیق پیش‌بینی جریان ورودی به سدها توسط مدل

رگرسیون بردار پشتیبان

مهدی عباسی^۱، شهاب عراقی‌نژاد^{۲*}، کیومرث ابراهیمی^۳

۱. فارغ‌التحصیل کارشناسی ارشد گروه مهندسی آبیاری و آبادانی، دانشکده مهندسی و فناوری کشاورزی، پردیس

کشاورزی و منابع طبیعی، دانشگاه تهران، کرج، ایران

۲. دانشیار، گروه مهندسی آبیاری و آبادانی، دانشکده مهندسی و فناوری کشاورزی، پردیس کشاورزی و منابع طبیعی،

دانشگاه تهران، کرج، ایران

۳. استاد، گروه مهندسی آبیاری و آبادانی، دانشکده مهندسی و فناوری کشاورزی، پردیس کشاورزی و منابع طبیعی،

دانشگاه تهران، کرج، ایران

(تاریخ دریافت: ۱۳۹۶/۱/۷ - تاریخ بازنگری: ۱۳۹۷/۴/۲ - تاریخ تصویب: ۱۳۹۷/۴/۱۲)

چکیده

پیش‌بینی دقیق هیدرولوژیکی یک ابزار کلیدی در برنامه‌ریزی‌های منابع آب است. از این‌رو در این مقاله با بهره‌گیری از مدل‌های رگرسیون بردار پشتیبان (SVR)، رگرسیون چند متغیره‌ی خطی (MLR) و خود همبسته‌ی میانگین متحرک (ARMA)، جریان ورودی به سدهای بختیاری و رودبار لرستان پیش‌بینی شده است. به منظور پیش‌پردازش داده‌های ورودی مدل‌ها از رویکرد میانگین متحرک استفاده شد. برای ارزیابی کارایی مدل‌ها از معیارهای ارزیابی نش - ساتکلیف (NSE)، جذر میانگین مربعات خطا (RMSE)، ضریب همبستگی (R) و دیاگرام تیلور استفاده شد. نتایج نشان داد که استفاده از روش پیش‌پردازش میانگین متحرک باعث بهبود عملکرد مدل‌های مورد استفاده شده است. بطوری که مقادیر NSE مربوط به مدل SVR با پیش‌پردازش میانگین متحرک در پیش‌بینی جریان ورودی به سدهای بختیاری و رودبار لرستان نسبت به مدل SVR بدون پیش‌پردازش به ترتیب ۱۳/۴ و ۶/۶ درصد بهبود داشته است.

واژه‌های کلیدی: پیش‌بینی سری زمانی، میانگین متحرک، سد بختیاری، سد رودبار، ARMA

مقدمه

امروزه با توجه به رشد جمعیت و افزایش بهداشت اجتماعی که به طور واضح متأثر از منابع آب می‌باشند، توانایی پیش‌بینی جریان برای مدیریت و برنامه‌ریزی منابع آب از اهمیت قابل توجهی برخوردار است. در دو ده اخیر بهره‌برداری اصولی از منابع آب به عنوان یک مسأله حائز اهمیت در مدیریت منابع آب به شمار می‌آید که این امر نیازمند مشخص بودن میزان آب در دسترس در افق آینده است که توسط پیش‌بینی‌های دقیق به منصفه ظهور می‌رسد.

روش‌های پیش‌پردازش در دو دسته مبتنی بر مدل‌ها (Model based methods) و مبتنی بر داده‌ها (Data driven methods) قرار می‌گیرند. داده‌های سری زمانی دارای سه مشکل اساسی از قبیل: (۱) بعضی داده‌ها دارای نویز سفید هستند. (۲) بعضی از مقادیر داده‌ها وجود ندارند و (۳) در بعضی

موارد هم ناسازگاری بین داده‌ها وجود دارد. این نواقص درون داده‌های حجم زیاد، بیشتر است و به همین خاطر توجه به آنها بسیار مهم است.

در این زمینه، Foroughi and Araghinejad (2017) به منظور پیش‌بینی جریان رودخانه‌های حوضه کرخه، از روش آماری تحلیل طیف تکین (SSA^۱) استفاده کردند. آنها در تحقیق خود نشان دادند که می‌توان با به‌کارگیری روش SSA مقادیر جریان رودخانه را با دقت مناسب پیش‌بینی نمود. Belayneh et al. (2014) از ترکیب تکنیک DWT با مدل‌های

ANN و SVR جهت بهبود پیش‌بینی بلند مدت خشکسالی با شاخص SPI بهره بردند. Yu et al. (2018) با استفاده از ترکیب پیش‌پردازش‌های FT^۲، EEMD^۳ و SSA و مدل SVR به مقایسه عملکرد این مدل‌های ترکیبی به منظور پیش‌بینی جریان ورودی به سد سه دره^۴ در چین پرداختند. آن‌ها نشان

1. Singular Spectrum Analysis
2. Fourier Transform
3. Ensemble Empirical Mode Decomposition
4. Three Gorges Dam

مدت کدورت جریان رودخانه (Nieto *et al.* 2014) و پیش‌بینی تقاضای آب در بخش‌های مختلف مصرف (Peña-Guzmán *et al.* 2016) را نام برد.

تحقیقات گسترده‌ای در زمینه مقایسه SVR با مدل‌های داده - مبنای دیگر در هیدرولوژی انجام شده است. به طوری که برتری این مدل را نسبت به سایر مدل‌های داده-مبنای خاص از قبیل MLR، K-MLR، نزدیکترین همسایه (K-NN) و برخی از ANN نشان می‌دهد. از همین رو در این میان بیشترین مقایسه مابین مدل ذکرشده و مدل ANN صورت گرفته است (Lin *et al.*, 2006; Wang *et al.*, 2010; Modaresi *et al.*, 2017; Kisi, 2015; Adnan *et al.*, 2017; Modaresi and Araghinejad, 2014).

(Seyam *et al.* 2017) از مدل SVR به منظور پیش‌بینی جریان رودخانه در مناطق مرطوب گرمسیری (مالزی) بهره گرفتند. آنها با تولید شش مدل مرتبط با SVR با تغییرات سطح آب و ارتفاع بارش در بالادست به پیش‌بینی جریان ساعتی رودخانه در مقطعی از سال ۲۰۱۱ در حوضه سلانگور پرداختند. نتایج آنها حاکی از کارایی SVR در پیش‌بینی جریان رودخانه بود.

در این مقاله به منظور پیش‌بینی جریان ورودی به سدهای بختیاری و رودبار لرستان به ارزیابی و استفاده از SVR، رگرسیون چند متغیره و مدل سری زمانی ARMA به همراه تکنیک پیش‌پردازش میانگین متحرک پرداخته شد. ارزیابی مدل‌های ترکیبی با مقایسه و انتخاب مدل برتر بر اساس بهترین عملکرد کارایی انجام شد. لذا رویکرد پیش‌پردازش میانگین متحرک در مدل‌های داده-مبنا به منظور پیش‌بینی جریان ورودی به سدهای بختیاری و رودبار لرستان بومی شد.

مواد و روش‌ها

منطقه مطالعاتی و داده‌ها

حوضه آبریز کارون با وسعت ۶۷،۲۵۷ کیلومترمربع و شبکه زهکشی مناسب به عنوان پرآب‌ترین حوضه در قسمت جنوب غربی ایران واقع شده است. با توجه به فعالیت توده‌های مدیترانه‌ای، ساختار زمین‌شناسی مناسب و دریافت زمستانه در بلندی‌های حوضه که به صورت برف انباشته شده است، موجب پدید آمدن رودخانه‌های متعدد با جریان دائمی در سطح آن گردیده است. در این مقاله به پیش‌بینی جریان ورودی به سدهای بختیاری و رودبار لرستان پرداخته شده است. اطلاعات

دادند که مدل ترکیبی FT-SVR نسبت به مدل‌های ترکیبی دیگر استفاده شده، عملکرد بهتری دارد.

Jamali and Ebrahimi (2010) با استفاده از مدل خود همبسته میانگین متحرک فصلی انباشته (SARIMA^۱)، قابلیت سری‌های زمانی را در پیش‌بینی پارامترهای هدایت الکتریکی و کلر در ایستگاه آستانه حوضه سفیدرود مورد بررسی قرار دادند. آنها نشان دادند که با اطمینان بالا می‌توان روش مدل‌سازی SARIMA را برای پیش‌بینی مقادیر هدایت الکتریکی و غلظت کلر در ایستگاه آستانه سفید، پیشنهاد کرد.

استفاده از شبکه‌های عصبی مصنوعی (ANN^۲) یکی از رایج‌ترین مدل‌ها برای پیش‌بینی‌های هیدرولوژیکی است. برخی مطالعات نشان می‌دهند ANN به علت سرعت پایین یادگیری و همگرایی، برای حل مسائل پیچیده در حیطه هیدرولوژی متمر ثمر واقع نمی‌شود (Guo *et al.* 2011). مدل ماشین بردار پشتیبان SVM^۳ معرفی شده توسط (Cortes and Vapnik, 1995) در سال‌های اخیر یکی از مؤثرترین ابزار پیش‌بینی به عنوان یک روش جایگزین برای ANN در نظر گرفته شده است (Guo *et al.* 2011). آموزش نسبتاً ساده مدل و عملکرد بالا برای داده‌هایی با ابعاد بزرگتر را می‌توان از مزایای مدل SVM به شمار آورد. وابستگی زیاد به تابع کرنل و مقدار پارامترهای مدل، از عیب‌های این مدل به شمار می‌آید. تعیین مقادیر پارامترهای مدل توسط الگوریتم‌های بهینه‌سازی برطرف می‌شوند.

Modaresi *et al.* (2017) به ارزیابی استراتژی‌های

میانگین‌گیری وزنی رتبه‌ای در ترکیب مدل‌های پیش‌بینی کننده جریان در رودخانه کرخه پرداختند. آنها نشان دادند که استراتژی وزن‌دهی متغیر دارای قابلیت بیشتری برای ارتقای نتایج پیش‌بینی نسبت به وزن‌دهی ثابت در مدل‌های بردار پشتیبان رگرسیونی (SVR^۴) است.

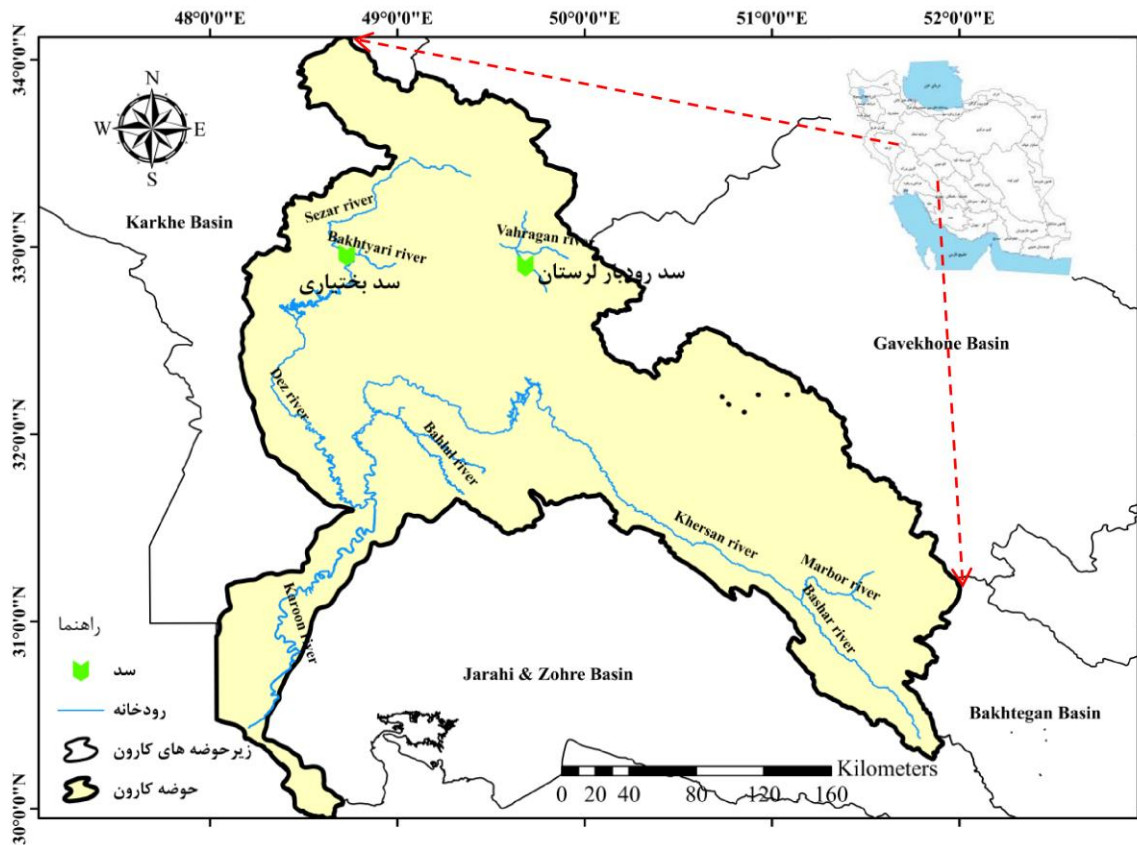
مدل SVR در زمینه‌های مختلفی به منظور مدل‌سازی ریاضی و عددی و پیش‌بینی شرایط متغیرهای وابسته در آینده، استفاده شده است که از این رو می‌توان به استفاده در مسئله مخزن به منظور شرایط خشکسالی (Chiang and Tsai, 2012)، پیش‌بینی جریان ورودی به مخزن در مقیاس‌های زمانی مختلف (Lin *et al.*, 2006)، پیش‌بینی سطح آب زیرزمینی در نزدیک محیط اشباع خاک مانند دریاچه (Gong *et al.* 2016)، تخمین تبخیر و تعرق روزانه مرجع (Wen *et al.* 2015)، پیش‌بینی بلند

1. Seasonal Autoregressive Integrated Moving Average
2. Artificial Neural Network
3. Support Vector Machine
4. Support Vector Regression

سدها و داده‌های به کار گرفته شده و موقعیت قرارگیری سدها، حوضه کارون به ترتیب در جدول (۱) و شکل (۱) آورده شده است. ایستگاه‌های هیدرومتری و شبکه‌ی زهکشی اصلی در سطح

جدول ۱. اطلاعات کلی جریان ماهانه و طول دوره آماری

ردیف	سد	ایستگاه‌های هیدرومتری	طول دوره آماری ایستگاه منتخب (ماه)	میانگین جریان ایستگاه (m^3/sec)	انحراف معیار m^3/sec
۱	بختیاری	تنگ پنج بختیاری مجموع دو ایستگاه	۲۰۱۵-۱۹۹۵	۱۴۱/۴	۱۲۱/۲۳
۲	رودبار لرستان	هیدرومتری زرد فهره، کاکلستان و قلیان	۲۰۱۵-۱۹۸۲	۲۳/۷	۲۷/۸۷



شکل ۱. موقعیت قرارگیری سدهای مورد مطالعه و شبکه زهکشی اصلی در حوضه کارون بزرگ

$$\text{Minimize } \frac{1}{2} \|w\|^2 + C \sum_{i=0}^l (\xi + \xi^*) \quad (\text{رابطه ۱})$$

$$\text{Subjected to } \begin{cases} y_i - (w, x_i) - b \leq \epsilon_i + \xi \\ (w, x_i) + b - y_i \leq \epsilon_i + \xi^* \\ \xi_i, \xi_i^* \geq 0 \quad i = 1, \dots, l \end{cases} \quad (\text{رابطه ۲})$$

ξ و ξ^* متغیرهای کمکی هستند که برای برخورد با قیود غیرممکن مسئله بهینه‌سازی، تعریف شده‌اند. ثابت C توازی مابین هموار کردن تابع (کاهش پیچیدگی مدل) و خطای تجربی را نشان می‌دهد. برای مثال اگر $C = \infty$ باشد، آنگاه

ماشین بردار پشتیبان SVM

Vapnik *et al.*, (1997) با معرفی تابع خطای جدید ϵ -insensitive نسخه جدیدی از SVM را که علاوه بر کلاس-بندی، قالب رگرسیونی این مدل که شامل ابرصفحه C و تابع کرنل است را ارائه نمودند و آن را بردار پشتیبان رگرسیونی (SVR) نامیدند. از SVR اغلب می‌توان به منظور پیش‌بینی بهره گرفت. فرض کنید مجموعه‌ی $S = \{(x_1, y_1), (x_2, y_2), \dots, (x_l, y_l)\}$ که در آن $x_i \in R^n$ مجموعه‌ی داده‌های آموزشی در فضای ورودی n بعدی است را داریم. نحوه بهینه این مسئله به فرم زیر است.

ارزیابی زیر استفاده شد:

نش-ساتکلیف

$$NS = 1 - \frac{\sum_{i=1}^n [Q_{obs,i} - Q_{sim,i}]^2}{\sum_{i=1}^n [Q_{obs,i} - Q_{obs}]^2} \quad (\text{رابطه ۴})$$

جذر میانگین مربعات خطا

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (Q_{obs,i} - Q_{sim,i})^2}{n}} \quad (\text{رابطه ۵})$$

ضریب همبستگی

$$r_{x,y} = \frac{\sum (Q_{obs,i} - \overline{Q_{obs}})(Q_{sim,i} - \overline{Q_{sim}})}{\sqrt{\sum (Q_{obs,i} - \overline{Q_{obs}})^2 \sum (Q_{sim,i} - \overline{Q_{sim}})^2}} \quad (\text{رابطه ۶})$$

که در اینجا $Q_{obs,i}$ و $Q_{sim,i}$ به ترتیب جریان مشاهده شده و جریان شبیه‌سازی شده در زمان i ام داده‌ها، Q_{obs} میانگین جریان مشاهده شده و در نهایت n تعداد داده‌های سری زمانی است.

محدوده تغییرات NSE از $(-\infty, 1)$ به ترتیب برای بهترین و بدترین عملکرد متغیر است. مقدار بین $(0, 1)$ سطح عملکرد قابل قبولی را توجیه می‌کند در صورتی که برای مقادیر $NSE \leq 0$ عملکرد غیرقابل قبول مدل را نشان می‌دهد. مقادیر کم RMSE در محدوده تغییر $(0, \infty)$ نشان از عملکرد بهتر مدل را داراست. ضریب همبستگی به منظور تعیین شدت همبستگی و وجود رابطه معنادار بین دو متغیر استفاده می‌شود.

آخرین معیار ارزیابی دیاگرام تیلور است. Taylor (2001) نمودار ریاضی را در سال ۱۹۹۴ ارائه داد و مقاله آن در سال ۲۰۰۱ به چاپ رسید. این دیاگرام با بهره‌گیری از سه پارامتر ارزیابی ضریب همبستگی، مجذور حداقل مربعات خطا و انحراف استاندارد به بررسی توانایی و عملکرد بهترین مدل، متغیر و پدیده در بین مدل‌ها، متغیرها و یا پدیده‌ها می‌پردازد.

نتایج و بحث

در این بخش با استفاده از مدل‌ها و معیارهای ارزیابی مذکور در بخش قبل، اقدام به ارزیابی و مقایسه عملکرد مدل‌ها در پیش‌بینی جریان ورودی به سدهای مورد مطالعه شده است. از این‌رو در ابتدا مدل‌های ترکیبی مختلفی تولید می‌شود. سپس با بهره‌گیری از معیارهای ارزیابی از میان مدل‌های ترکیبی تولیدشده و مدل سری زمانی ARMA به انتخاب برترین مدل به منظور پیش‌بینی جریان ورودی به سدهای بختیاری و رودبار لرستان پرداخته شد. برای این امر از میانگین‌های متحرک دو ماهه و سه ماهه (SMA2) و (SMA3) داده‌های جریان به عنوان ورودی به مدل‌های داده-مبنای ذکرشده استفاده شده است. مشخصات

هدف، کمینه‌کردن مخاطره تجربی^۱ بدون در نظر گرفتن پیچیدگی‌های مدل خواهد بود. ϵ میزان انحراف مابین مقادیر هدف واقعی و نمونه‌های آموزشی است. به عبارتی دیگر می‌توان گفت که خطاهای کمتر از ϵ را نادیده گرفته اما خطاهای بزرگتر از آن پذیرفتنی نیست که متغیرهای کمکی به این منظور تعریف می‌شوند.

با توجه به اینکه داده‌های هیدرولوژی ماهیتی غیرخطی دارند، برای کاهش بعد و استفاده از شکل استاندارد SVR به کارگیری تابع‌های کرنل نیاز است. تحقیقات بی‌شمار انجام شده در مقوله‌ی پیش‌بینی جریان، عملکرد بالای تابع کرنل شعاعی پایه (RBF^T) را برای SVM نشان می‌دهند (Yu et Sivapragasam et al., 2001; Choy and Chan., 2003; al., 2004)؛ بنابراین تابع کرنلی که در این تحقیق به کار گرفته شده است، تابع کرنل $RBF = \exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right)$ بوده است. شایان ذکر است که دقت و تأثیر مدل SVM به منظور پیش‌بینی، به شدت تحت تأثیر پارامترهای مدل γ, ϵ and C می‌باشد. از این‌رو، در این تحقیق به منظور بهینه‌سازی این پارامترها، از تکنیک گریدبندی بهره گرفته شده است. برای این منظور با کدنویسی تابعی در محیط R studio و کمینه‌کردن مقدار مربعات خطای تابع تولیدشده، اقدام به یافتن بهترین مقادیر پارامترهای مدل SVR شده است.

سری زمانی خود همبسته میانگین متحرک ARMA

مدل خود همبسته میانگین متحرک (ARMA) مدلی ترکیبی از مدل‌های AR و MA است. در این مدل متغیر Z_t علاوه بر مقادیر قبلی خود به مقادیر قبلی ϵ_t نیز وابستگی دارد. فرم پایه این مدل برای متغیر نرمال استانداردشده Z_t به صورت رابطه ۳ می‌باشد.

$$\begin{aligned} Z_t &= \phi_1 Z_{t-1} + \phi_2 Z_{t-2} + \dots + \phi_p Z_{t-p} + \epsilon_t - \\ &\theta_1 \epsilon_{t-1} - \dots - \theta_q \epsilon_{t-q} = \sum_{i=1}^p \phi_i Z_{t-i} - \sum_{j=0}^q \theta_j \epsilon_{t-j}, \quad \text{for } \theta_0 = 1 \end{aligned} \quad (\text{رابطه ۳})$$

در اینجا Z_t متغیر نرمال استاندارد شده وابسته به زمان، θ_j پارامتر MA ، ϕ_j پارامتر AR از مدل MA و ϵ_t متغیر تصادفی مستقل از زمان (اغتشاشات خالص) نام دارند.

معیارهای ارزیابی

برای ارزیابی و مقایسه مدل‌های ترکیبی ایجاد شده به منظور پیش‌بینی جریان ورودی به سدها در این مطالعه از معیارهای

1. Empirical risk
2. Radial Basin Function

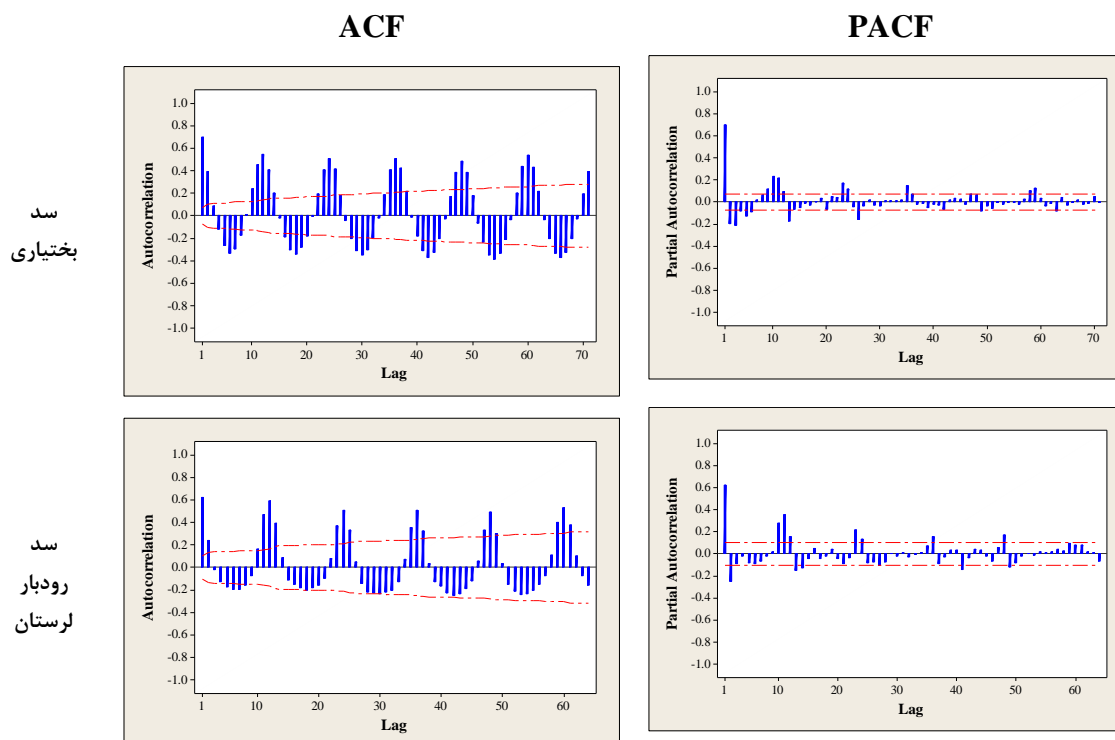
ARIMA تفاوت چندانی را در نتایج به دنبال نخواهد داشت. مرتبه‌هایی که برای مدل ARMA به منظور مدل‌سازی جریان ورودی به سدهای بختیاری و رودبار لرستان می‌توان استفاده کرد، در جدول (۳) آورده شده است. همچنین لازم به ذکر است که مرتبه مدل AR(p) با کمک گرفتن از نمودار PACF تعیین و مرتبه قسمت MA(q) با بهره‌گیری از نمودار ACF محاسبه و تعیین می‌گردد. در نهایت با توجه به نمودارهای ACF، PACF و تحلیل‌های دیگر نظیر معیار اطلاعات آکائیک (Akaike, 1974)، بهترین مدل سری زمانی جریان ورودی به سدهای بختیاری و رودبار لرستان، مدل ARMA(2,1) انتخاب شده است.

چهار مدل ترکیبی و مدل سری زمانی ARMA به همراه علامت اختصاری در نظر گرفته شده برای آنها در جدول (۲) ارائه شده است.

مرتبه‌های مدل ARMA(p,q): p: مرتبه مدل خود همبسته (AR) q: مرتبه مدل میانگین محرک (MA) را می‌توان توسط تابع خود همبسته (ACF) و تابع خود همبسته جزئی (PACF) تعیین نمود. طبق شکل (۲) که نمودارهای ACF و PACF جریان ورودی به سدهای بختیاری و رودبار لرستان را ارائه می‌دهند، نشان از میرایی نمودارهای ACF و PACF می‌باشد. از همین رو به‌کارگیری مدل ARMA و

جدول ۲. شماره‌گذاری مدل‌ها به کار گرفته شده به اختصار

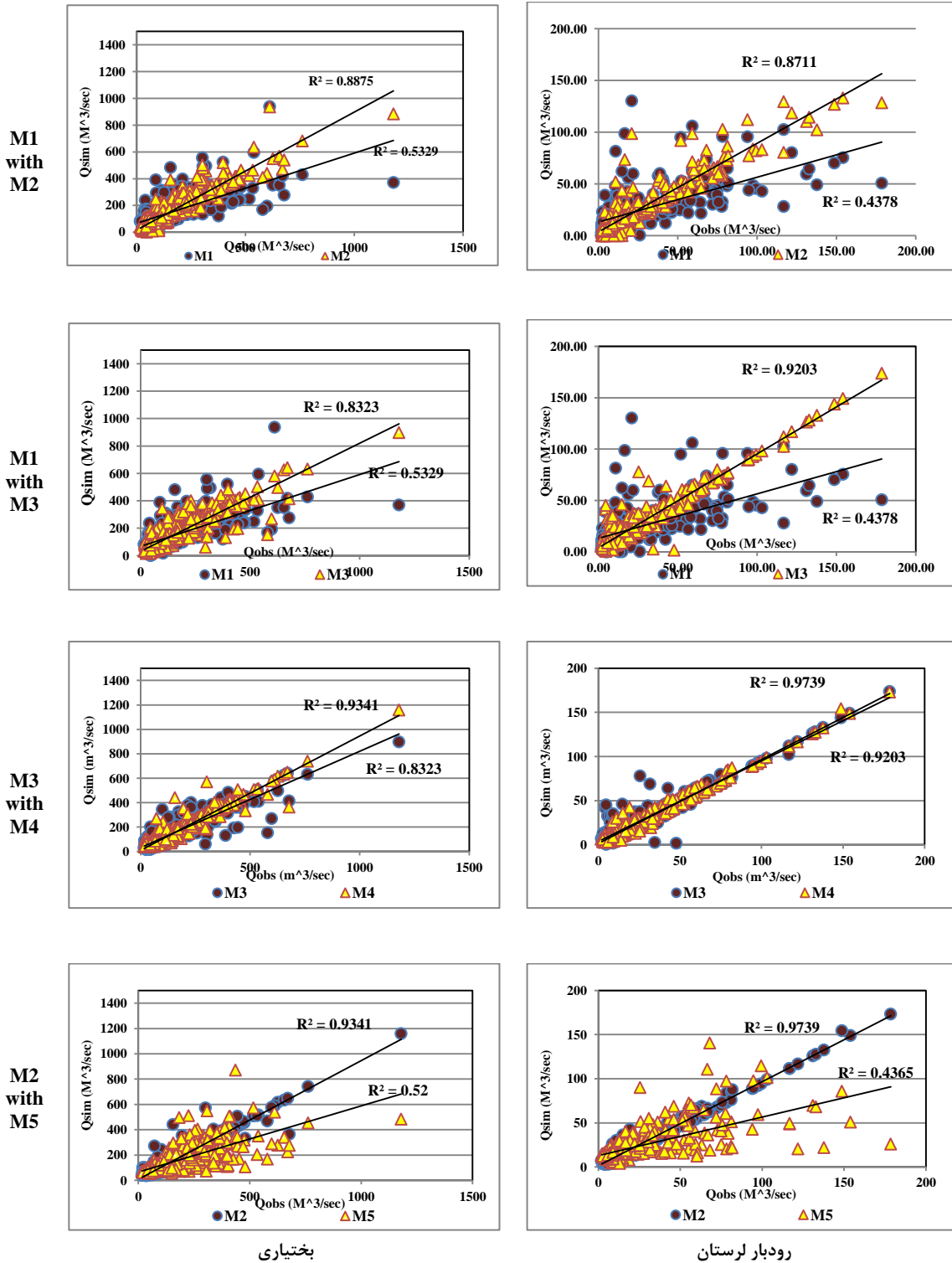
ردیف	مدل	شماره مدل	Model
۱	مدل رگرسیون چند متغیره خطی با گام‌های تأخیر زمانی یک، دو و سه‌ماهه	۱	M1
۲	مدل رگرسیون چند متغیره خطی با (SMA2) و (SMA3)	۲	M2
۳	مدل بردار پشتیبان رگرسیونی با گام‌های تأخیر زمانی یک، دو و سه‌ماهه	۳	M3
۴	مدل بردار پشتیبان رگرسیونی با (SMA2) و (SMA3)	۴	M4
۵	مدل سری زمانی ARMA	۵	M5



شکل ۲. ACF و PACF جریان ورودی به سدهای بختیاری و رودبار لرستان

جدول ۳. مرتبه‌های مورد استفاده در مدل ARMA به منظور پیش‌بینی جریان ورودی به سدهای مورد مطالعه

پارامترهای ARMA	سد	رودبار لرستان
p	{۰,۱,۲,۳}	{۰,۱,۲,۳,۴,۵}
q	{۰,۱,۲}	{۰,۱,۲}



شکل ۳. مقایسه نمودار نقطه‌ای مشاهده شده (محور افقی) و شبیه‌سازی شده (محور عمودی) مدل‌های ترکیبی تولیدشده برای جریان ورودی به سدهای بختیاری و رودبار لرستان

مدل SVR را نسبت به مدل MLR در شرایط یکسان ورودی‌ها نشان می‌دهد. به طوری که معیار ارزیابی ضریب همبستگی به ترتیب برای داده‌های جریان ورودی به سدهای بختیاری و رودبار لرستان افزایش ۰/۱۸ و ۰/۳۰ را به همراه داشته است. در ردیف سوم، مدل SVR در شرایطی که متغیرهای ورودی غیر یکسان باشند، مورد مقایسه قرار گرفت. این مقایسه برتری متغیرهای ورودی به SMA2 و SMA3 را نسبت به گام‌های تأخیر زمانی یک، دو و سه‌ماهه در مدل SVR نشان می‌دهد. در ردیف انتهایی بهترین مدل سری زمانی ARMA برای جریان ورودی به سدهای بختیاری و رودبار لرستان با بهترین مدل ترکیبی انتخاب شده، مورد مقایسه قرار داده شده است. این مقایسه نشان از برتری مدل ترکیبی M4 نسبت به بهترین مدل سری زمانی ARMA برای جریان ورودی به سدهای مورد مطالعه را دارد. همچنین مقادیر معیارهای ارزیابی ضریب همبستگی، RMSE و NSE برای مدل‌های مختلف به منظور پیش‌بینی جریان ورودی به سدهای بختیاری و رودبار لرستان به ترتیب در جدول‌های (۴) و (۵) ارائه شده است. همچنین سری زمانی جریان مشاهداتی و شبیه‌سازی شده مدل‌های ترکیبی M1 & M2، M1 & M3 و M4 & M5 به ترتیب در شکل‌های (۴)، (۵) و (۶) به تصویر کشیده شده است.

در شکل (۳) با بهره‌گیری از نمودار نقطه‌ای به مقایسه زوج مرتب مدل‌های ذکر شده در جدول (۲) برای داده‌های جریان ورودی به سدهای مورد مطالعه در این تحقیق پرداخته شده است. در این شکل در ردیف اول مقایسه مدل‌های ترکیبی M1&M2 به تصویر کشیده شده است، سپس به ترتیب در ردیف‌های بعد مقایسه مابین مدل‌های ترکیبی M1&M3، M3&M4 و M4&M5 صورت پذیرفته است.

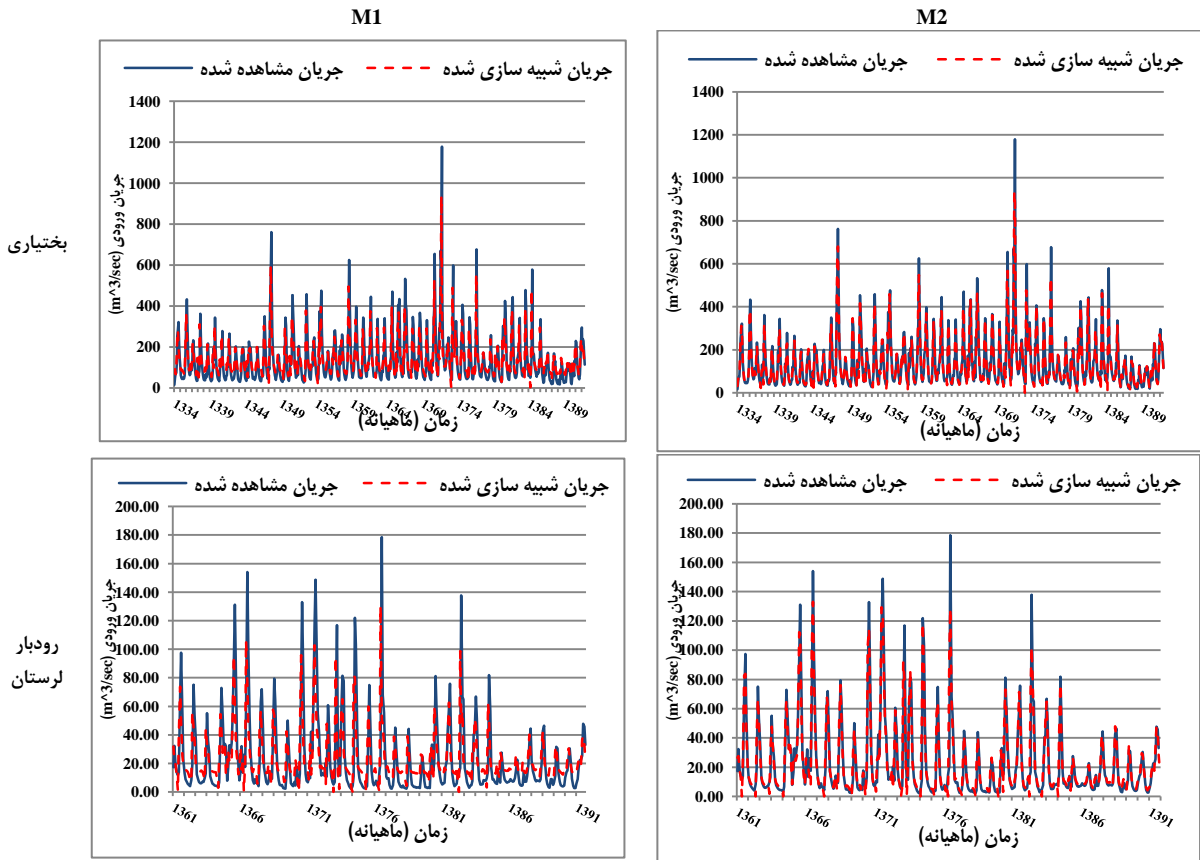
در ردیف اول شکل (۳) که به مقایسه جریان مشاهداتی و شبیه‌سازی شده مدل‌های ترکیبی M1&M2 پرداخته شده است، افزایش ۰/۲۱ معیار ارزیابی ضریب همبستگی برای داده‌های جریان ورودی به سد بختیاری و افزایش ۰/۲۷ معیار برای داده‌های جریان ورودی به سد رودبار لرستان حاکی از برتری عملکرد مدل M2 نسبت به M1 دارد که این امر اهمیت بالای انتخاب متغیرهای مناسب به عنوان ورودی به مدل MLR را نشان می‌دهد. به طوری که مدل M2 با استفاده از میانگین متحرک‌های دو و سه‌ماهه، افزایش چشم‌گیری در شبیه‌سازی جریان ورودی به سدهای مورد مطالعه نسبت به مدل M1 با متغیرهای ورودی گام‌های زمانی تأخیر یک، دو و سه‌ماهه را به همراه دارد. در ردیف دوم شکل (۳) با ارائه مدل‌های ترکیبی M1&M3 سعی در مقایسه مدل‌های بردار پشتیبان رگرسیونی و MLR با متغیر ورودی یکسان شده است. نتایج این ردیف برتری

جدول ۴. مقادیر معیارهای ارزیابی برای مدل‌های مختلف بمنظور پیش‌بینی جریان ورودی به سد بختیاری

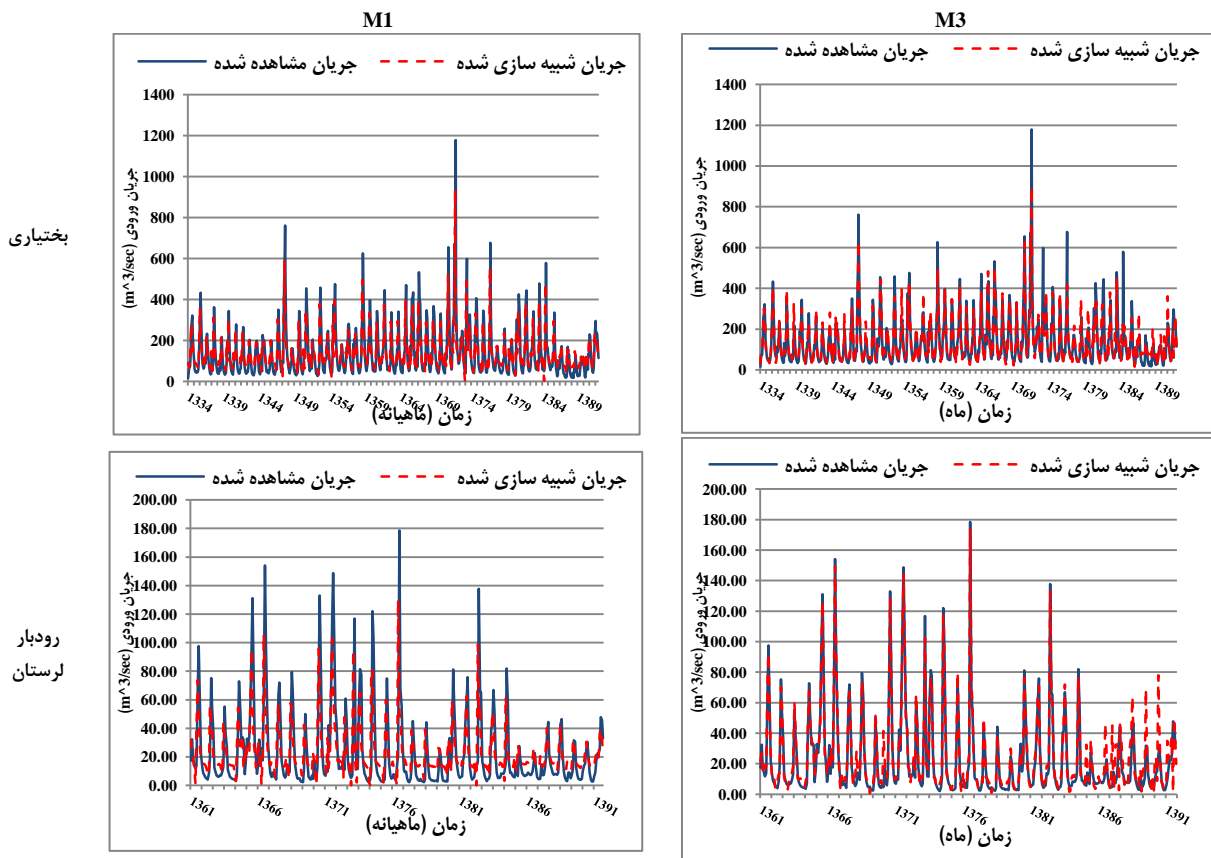
ردیف	مدل	R	RMSE (m ³ /sec)	NSE
۱	مدل رگرسیون چند متغیره خطی با گام‌های تأخیر زمانی یک، دو و سه‌ماهه	۰/۷۳	۸۲/۹۰	۰/۵۲
۲	مدل رگرسیون چند متغیره خطی با (SMA2) و (SMA3)	۰/۹۴	۴۲/۷۰	۰/۸۸
۳	مدل بردار پشتیبان رگرسیونی با گام‌های تأخیر زمانی یک، دو و سه‌ماهه	۰/۹۱	۴۹/۷۵	۰/۸۲
۴	مدل بردار پشتیبان رگرسیونی با (SMA2) و (SMA3)	۰/۹۷	۳۱/۱۴	۰/۹۳
۵	مدل سری زمانی ARMA	۰/۷۲	۸۴/۰۲	۰/۵۲

جدول ۵. مقادیر معیارهای ارزیابی برای مدل‌های مختلف بمنظور پیش‌بینی جریان ورودی به سد رودبار لرستان

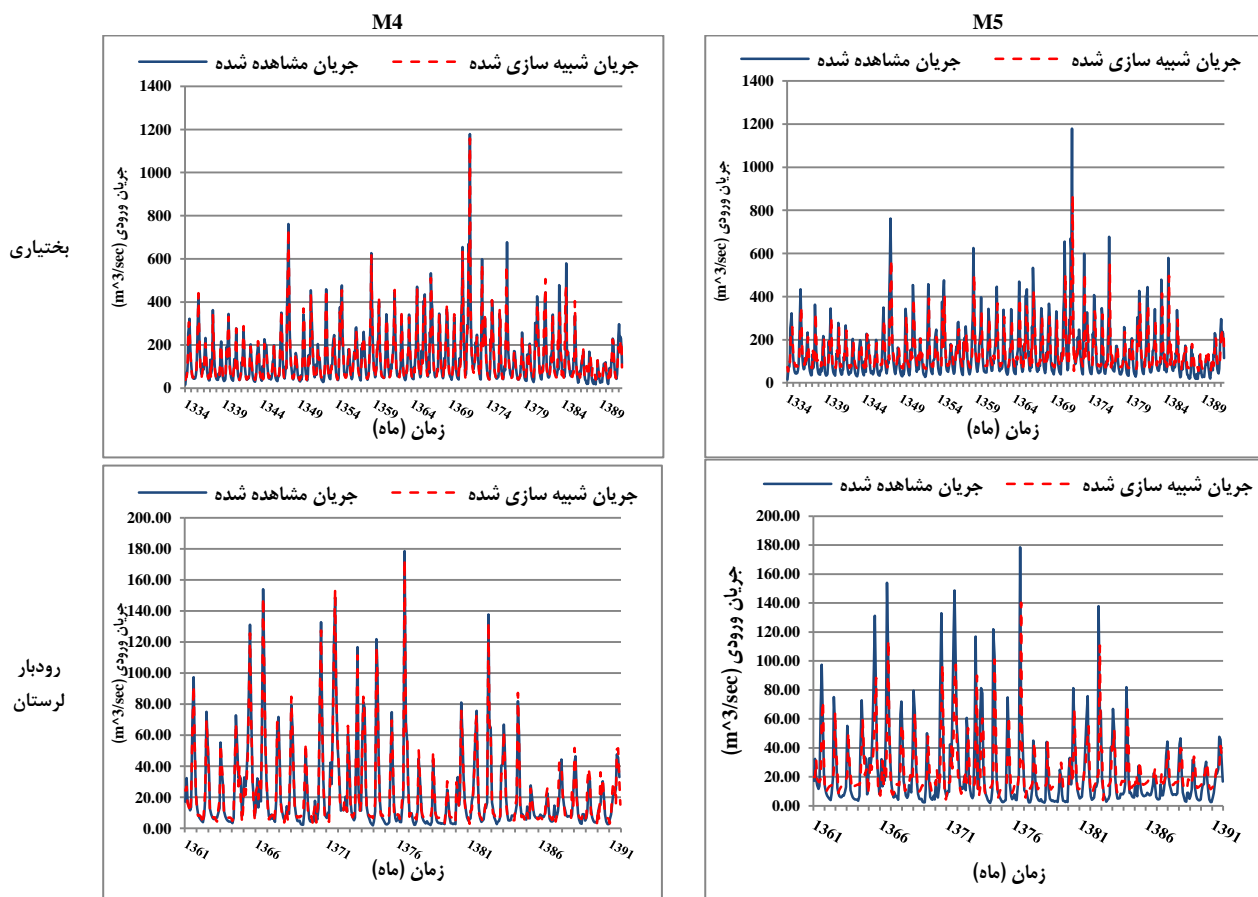
ردیف	مدل	R	RMSE (m ³ /sec)	NSE
۱	مدل رگرسیون چند متغیره خطی با گام‌های تأخیر زمانی یک، دو و سه‌ماهه	۰/۶۶	۲۱	۰/۴۳
۲	مدل رگرسیون چند متغیره خطی با (SMA2) و (SMA3)	۰/۹۳	۱۰/۳۷	۰/۸۶
۳	مدل بردار پشتیبان رگرسیونی با گام‌های تأخیر زمانی یک، دو و سه‌ماهه	۰/۹۶	۸/۹۰	۰/۹۱
۴	مدل بردار پشتیبان رگرسیونی با (SMA2) و (SMA3)	۰/۹۸	۵/۳۱	۰/۹۷
۵	مدل سری زمانی ARMA	۰/۶۶	۲۰/۹۳	۰/۴۴



شکل ۴. نمودار سری زمانی مقایسه‌ی جریان ورودی مشاهداتی و شبیه‌سازی شده برای مدل‌های ترکیبی M1 و M2



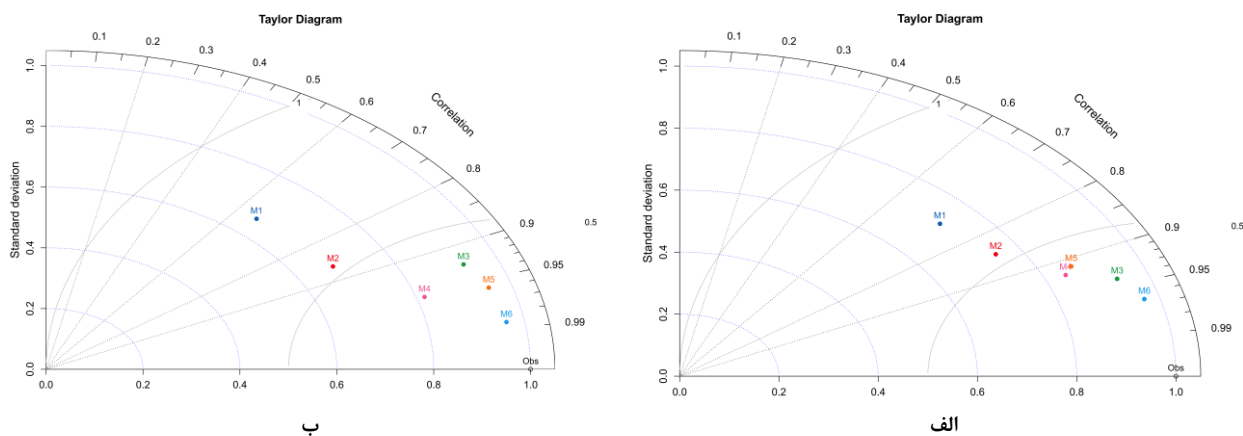
شکل ۵. نمودار سری زمانی مقایسه‌ی جریان ورودی مشاهداتی و شبیه‌سازی شده برای مدل‌های ترکیبی M1 و M3



شکل ۶. نمودار سری زمانی مقایسه‌ی جریان ورودی مشاهداتی و شبیه‌سازی شده برای مدل‌های ترکیبی M4 و M5

مقادیر منفی را برای جریان ورودی به سدهای مورد مطالعه شبیه‌سازی می‌کند که این نکته برای پیش‌بینی جریان بسیار حائز اهمیت بوده و به هیچ وجه قابل چشم‌پوشی نیست؛ بنابراین با در نظر گرفتن قید بزرگتر از صفر بودن داده‌های جریان، مقادیر منفی شبیه‌سازی شده تبدیل به صفر می‌شود.

در شکل (۷) از دیاگرام تیلور به منظور ارزیابی چهار مدل ترکیبی تولیدشده و بهترین مدل سری زمانی ARMA استفاده شد. این دیاگرام به صورت بصری برتری عملکرد مدل ترکیبی M4 را نشان می‌دهد. مدل ترکیبی M2 بعد از مدل ترکیبی M4 که به عنوان مدل برتر شناخته‌شده، قرار می‌گیرد. با توجه به ماهیت مثبت بودن داده‌های جریان رودخانه، مدل ترکیبی M2



شکل ۷. دیاگرام تیلور شش مدل ترکیبی تولیدشده برای جریان ورودی به سدهای الف) بختباری ب) رودبار لرستان

تاریخی جریان، در گام‌های زمانی یک، دو و سه‌ماهه به عنوان متغیر ورودی به مدل‌های مذکور بهره گرفته شد و منجر به تولید چهار مدل ترکیبی M1، M2، M3 و M4 به همراه بهترین مدل سری زمانی ARMA (M5) گردید. نتایج این تحقیق حاکی از آن است که مدل ترکیبی M4 با مقادیر ضریب NSE ۰/۹۳ و ۰/۹۷ به ترتیب برای شبیه‌سازی داده‌های جریان ورودی به سدهای بختیاری و رودبار لرستان دارای بهترین عملکرد بوده است. در مقابل مدل M1 نتایج مطلوبی ارائه نداده است و در بین تمامی مدل‌های ذکر شده ضعیف‌ترین عملکرد را به خود اختصاص داده است. نکته حائز اهمیت دیگری که باید به آن اشاره کرد، مقادیر منفی شبیه‌سازی شده برای جریان رودخانه که طبیعتاً مثبت است، توسط مدل‌های ترکیبی M1 و M2 است که این امر ضعف بالای MLR را در شبیه‌سازی نشان می‌دهد. همچنین با وجود مقادیر مناسب معیارهای ارزیابی برای مدل ترکیبی M2، این مدل مقادیر منفی را برای جریان ورودی به سدهای مورد مطالعه شبیه‌سازی می‌کند که با در نظر گرفتن قید بزرگتر از صفر بودن داده‌های جریان، مقادیر منفی شبیه‌سازی شده تبدیل به صفر می‌شود. در پایان، موضوع قابل توجه تأثیر انتخاب مناسب متغیرهای ورودی به مدل‌هاست به طوری که استفاده از تکنیک میانگین متحرک نقش بسزایی در بهبود مدل‌های MLR و SVR به همراه داشته است. برای مثال معیار ارزیابی NSE برای مدل MLR به ترتیب در سدهای بختیاری و رودبار لرستان از ۰/۵۲ به ۰/۸۸ و از ۰/۴۳ به ۰/۸۶ افزایش چشم‌گیری را نشان می‌دهد.

سپاسگزاری

بدین وسیله از دانشگاه تهران به خاطر تأمین امکانات لازم برای انجام این تحقیق و تهیه مقالات مربوطه تشکر و قدردانی می‌شود.

همان‌طور که در شکل (۷) مشاهده می‌شود کمترین همبستگی بین داده‌های مشاهداتی و شبیه‌سازی جریان ورودی به سدها مربوط به مدل ترکیبی M1 است که به ترتیب برای سدهای بختیاری و رودبار لرستان، مقادیر ضریب همبستگی ۰/۷۳ و ۰/۶۶ را بدست آورده است. در صورتی که بهترین و بیشترین همبستگی مربوط به مدل ترکیبی M4 با مقادیر ضریب همبستگی ۰/۹۷ و ۰/۹۸ می‌باشد.

نتایج مقاله حاضر در ارزیابی رویکرد پیش‌پردازش میانگین متحرک با نتایج تحقیقات انجام شده قبلی در این زمینه در سطح جهانی، نشان از بهبود عملکرد مدل SVR با رویکرد ارائه شده دارد. (Adnan et al. (2017 با بهره‌گیری از مدل SVR مقدار ضریب همبستگی را برای داده‌های جریان ماهانه رودخانه دانیوب^۱ مقدار ۰/۸۶ بدست آوردند. Yu et al. (2018) ترکیب FT-SVR را به عنوان بهترین مدل برای پیش‌بینی جریان ماهانه به سد یانگ ته‌سه نشان دادند. در نهایت، در این مقاله با رویکرد پیش‌پردازش میانگین متحرک، مقدار مدل SVR با رویکرد مورد ذکر (M4) برای سدهای بختیاری و رودبار لرستان به ترتیب مقادیر ۰/۹۷ و ۰/۹۸ را نشان داد.

نتیجه‌گیری

نظر به اینکه سدها به عنوان یکی از منابع تأمین نیازها و مصارف آبی در حیطه‌ی مدیریت و برنامه‌ریزی منابع آب به شمار می‌آید، لذا پیش‌بینی جریان ورودی به سدها از اهمیت بسزایی برخوردار است. در این تحقیق با استفاده از مدل‌های داده-مبنای رگرسیون چند متغیره، بردار پشتیبان رگرسیونی و مدل سری زمانی ARMA اقدام به پیش‌بینی سری زمانی جریان ورودی به سدهای بختیاری و رودبار لرستان شده است. از همین رو از تکنیک میانگین متحرک و گام‌های زمانی تأخیر مقادیر

1. Dainyor

REFERENCE

- Belayneh, A., Adamowski, J., Khalil, B. and Ozga-Zielinski, B. (2014). "Long-term SPI drought forecasting in the Awash River Basin in Ethiopia using wavelet neural network and wavelet support vector regression models," *Journal of Hydrology*, vol. 508, pp. 418-429.
- Adnan, R. M., Yuan, X., Kisi, O., & Yuan, Y. (2017). Streamflow Forecasting Using Artificial Neural Network and Support Vector Machine Models. *American Scientific Research Journal for Engineering, Technology, and Sciences (ASRJETS)*, 29(1), 286-294.
- Akaike, H. (1974). A new look at the statistical model identification. *IEEE transactions on automatic control*, 19(6), 716-723.
- Chiang, J. L., & Tsai, Y. S. (2012). Reservoir drought prediction using support vector machines. *In Applied Mechanics and Materials* (Vol. 145, pp. 455-459). Trans Tech Publications.
- Choy, K. Y., & Chan, C. W. (2003). Modelling of river discharges and rainfall using radial basis function networks based on support vector regression. *International Journal of Systems Science*, 34(14-15), 763-773.
- Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine learning*, 20(3), 273-297.
- Foroughi, F., Araghinejad, S. (2017). Long-lead streamflow forecasting using singular spectrum analysis in the Karkheh basin. *Iranian Journal of Soil and Water Research*, 48(2), 309-321. doi:

- 10.22059/ijswr.2017.62633. (in Farsi)
- Gong, Y., Zhang, Y., Lan, S., & Wang, H. (2016). A comparative study of artificial neural networks, support vector machines and adaptive neuro fuzzy inference system for forecasting groundwater levels near Lake Okeechobee, Florida. *Water Resources Management*, 30(1), 375-391.
- Guo, J., Zhou, J., Qin, H., Zou, Q., & Li, Q. (2011). Monthly streamflow forecasting based on improved support vector machine model. *Expert Systems with Applications*, 38(10), 13073-13081.
- Jamali, B., Ebrahimi, K. (2010). water quality time series forecasting using linear models random case study: Sefid Rud river. *Agricultural Engineering Research*, 12 (3), 31-44. doi: 10.22092/jaer.2011.100317. (in Farsi)
- Kisi, O. (2015). Streamflow forecasting and estimation using least square support vector regression and adaptive neuro-fuzzy embedded fuzzy c-means clustering. *Water resources management*, 29(14), 5109-5127.
- Lin, J. Y., Cheng, C. T., & Chau, K. W. (2006). Using support vector machines for long-term discharge prediction. *Hydrological Sciences Journal*, 51(4), 599-612.
- Modaresi F., Araghinejad S., Ebrahimi K. (2017a). Assessment of Ordered Weighted Averaging Strategies in Combination of Streamflow Forecasting Models. *jwmseir*. 10 (35):15-25. URL: <http://jwmsei.ir/article-1-469-fa.html> (in Farsi)
- Modaresi, F., Araghinejad, S., & Ebrahimi, K. (2017b). A Comparative Assessment of Artificial Neural Network, Generalized Regression Neural Network, Least-Square Support Vector Regression, and K-Nearest Neighbor Regression for Monthly Streamflow Forecasting in Linear and Nonlinear Conditions. *Water Resources Management*, 1-16.
- Modaresi, F., & Araghinejad, S. (2014). A comparative assessment of support vector machines, probabilistic neural networks, and K-nearest neighbor algorithms for water quality classification. *Water resources management*, 28(12), 4095-4111.
- Nieto, P. G., García-Gonzalo, E., Fernández, J. A., & Muñoz, C. D. (2014). Hybrid PSO–SVM-based method for long-term forecasting of turbidity in the Nalón river basin: A case study in Northern Spain. *Ecological Engineering*, 73, 192-200
- Peña-Guzmán, C., Melgarejo, J., & Prats, D. (2016). Forecasting Water Demand in Residential, Commercial, and Industrial Zones in Bogotá, Colombia, Using Least-Squares Support Vector Machines. *Mathematical Problems in Engineering*, 2016.
- Seyam, M., Othman, F., & El-Shafie, A. (2017). Prediction of Stream Flow in Humid Tropical Rivers by Support Vector Machines. *In MATEC Web of Conferences* (Vol. 111, p. 01007). EDP Sciences.
- Sivapragasam, C., Liong, S. Y., & Pasha, M. F. K. (2001). Rainfall and runoff forecasting with SSA–SVM approach. *Journal of Hydroinformatics*, 3(3), 141-152.
- Taylor, K. E. (2001). Summarizing multiple aspects of model performance in a single diagram. *Journal of Geophysical Research: Atmospheres*, 106(D7), 7183-7192.
- Vapnik, V., Golowich, S. E., & Smola, A. J. (1997). Support vector method for function approximation, regression estimation and signal processing. In *Advances in neural information processing systems* (pp. 281-287).
- Wang, W., Nie, X., & Qiu, L. (2010, October). Support vector machine with particle swarm optimization for reservoir annual inflow forecasting. In *Artificial Intelligence and Computational Intelligence (AICI), 2010 International Conference on* (Vol. 1, pp. 184-188). IEEE.
- Wen, X., Si, J., He, Z., Wu, J., Shao, H., & Yu, H. (2015). Support-vector-machine-based models for modeling daily reference evapotranspiration with limited climatic data in extreme arid regions. *Water Resources Management*, 29(9), 3195-3209.
- Yu, X., Liong, S. Y., & Babovic, V. (2004). EC-SVM approach for real-time hydrologic forecasting. *Journal of Hydroinformatics*, 6(3), 209-223.
- Yu, X., Zhang, X., & Qin, H. (2018). A data-driven model based on Fourier transform and support vector regression for monthly reservoir inflow forecasting. *Journal of Hydro-environment Research*, 18, 12-24.