Archive of SID

نشریه علمی-پژوهشی مهندسی و مدیریت آبخیز
جلد ۱۱، شماره ۴، ۱۳۹۸، صفحات ۱۰۴۴-۱۰۵۸

Journal of Watershed Engineering and Management
Volume 11, Issue 4, 2019, Pages 1044-1058
DOI: 10.22092/ijwmse.2018.108926.1251

پیش‌بینی دبی ماهانه ورودی به سد بوستان در استان گلستان با استفاده از مدل‌های داده‌کاوی و ترکیبی

میثم صمدی*[1]، عبدالرضا بهره‌مند[2] و ابوالحسن فتح‌آبادی[3]

[1] دانشجوی دکتری علوم و مهندسی آبخیزداری، دانشکده مرتع و آبخیزداری، دانشگاه علوم کشاورزی و منابع طبیعی گرگان، [2] دانشیار، دانشکده مرتع و آبخیزداری، دانشگاه علوم کشاورزی و منابع طبیعی گرگان و [3] استادیار، دانشکده کشاورزی و منابع طبیعی، دانشگاه گنبد کاووس

تاریخ دریافت: ۱۳۹۶/۱۰/۰۷ تاریخ پذیرش: ۱۳۹۷/۰۹/۱۰

چکیده

در هر برنامه مدیریتی برای منابع آب، آگاهی از شرایط آینده به‌منظور تخصیص بهینه منابع آب به بخش‌های مختلف از قبیل شرب، کشاورزی و غیره لازم می‌باشد. آن‌چه در این میان مهم می‌باشد، پیش‌بینی مقادیر جریان ورودی به سیستم منابع آب در ماه‌های آینده است. در این راستا، استفاده از روش‌هایی که بتواند با کمینه خطا و با توجه به داده و اطلاعات موجود، جریان رودخانه را پیش‌بینی کند، از اهمیت فراوانی برخوردار می‌باشد. در پژوهش حاضر، مقادیر دبی ماهانه ورودی به سد بوستان برای آینده با استفاده از داده‌های هیدرومتری ایستگاه تمر و به‌کارگیری سه مدل سری زمانی، شبکه عصبی مصنوعی و ماشین بردار پشتیبان و همچنین، سه مدل ترکیبی پیش‌بینی شد. سپس، با استفاده از معیارهای ارزیابی اقدام به مقایسه عملکرد هر کدام از مدل‌ها شد. با توجه به نتایج به‌دست آمده در مدل، سری زمانی بر اساس کمینه بودن معیارهای آکاییک و شوارتز، مدل ARIMA (2,0,0) (1,0,1) به‌عنوان مدل برتر انتخاب شد. در مدل شبکه عصبی، شبکه با ورودی ۲ و ۴ نرون و در مدل SVM شبکه با ورودی ۳، به‌عنوان شبکه برتر انتخاب شدند. در نهایت، با توجه به نتایج به‌دست آمده از معیارهای ارزیابی، مدل سری زمانی بهترین عملکرد را داشته است که مقادیر معیارهای میانگین مربعات خطا، متوسط مقادیر مطلق خطای نسبی، میانگین مطلق خطا و نش-ساتکلیف برای این مدل به‌ترتیب برابر با ۰/۸۸، ۴/۷۱، ۰/۰۲۴- و ۰/۳۶ به‌دست آمد. در نتیجه، مدل سری زمانی به‌عنوان بهترین مدل برای پیش‌بینی دبی ماهانه در این ایستگاه معرفی شد.

واژه‌های کلیدی: ایستگاه تمر، مدل سری زمانی، مدل شبکه عصبی مصنوعی، مدل ماشین بردار پشتیبان، منابع آب

مقدمه

با توجه به محدودیت منابع آبی در طبیعت، توزیع زمانی و مکانی غیر یکنواخت منابع آبی، افزایش آلودگی‌ها، تخریب منابع طبیعی و همچنین، افزایش جمعیت، رشد و توسعه جوامع شهری و فعالیت‌های کشاورزی و صنعتی ضرورت برنامه‌ریزی و مدیریت صحیح منابع آبی اجتناب‌ناپذیر است. در هر برنامه مدیریتی، برای منابع آب آگاهی از شرایط آینده به‌منظور تخصیص بهینه منابع آب به بخش‌های مختلف از قبیل شرب، کشاورزی و غیره لازم می‌باشد. در یک سیستم منابع آب اغلب نیاز بخش‌های مختلف از قبیل کشاورزی، شرب و غیره تغییر چندانی نکرده و

* نشانی مکاتبات: meisamadi68@gmail.com

www.SID.ir

Archive of SID

به‌راحتی قابل پیش‌بینی می‌باشد. آن‌چه در این میان مهم می‌باشد، پیش‌بینی مقادیر جریان ورودی به سیستم منابع آب در ماه‌های آینده می‌باشد. به‌منظور پیش‌بینی جریان رودخانه در طی سالیان متمادی روش‌های مختلفی ابداع شده است. در این میان، روش سری‌های زمانی در طی چهار دهه اخیر به‌طور گسترده‌ای برای پیش‌بینی جریان رودخانه مورد استفاده قرار گرفته است. امروزه تجزیه و تحلیل سری‌های زمانی به‌طور وسیعی در بسیاری از شاخه‌های مهندسی، علوم فیزیک و اقتصاد مورد استفاده واقع می‌شود و می‌توان گفت که بیشتر شاخه‌های علوم منجر به مطالعه داده‌هایی می‌شوند که به‌شکل سری‌های زمانی رخ می‌دهند (Niroumand و Bozorgnia، ۱۹۹۳). سری زمانی مجموعه‌ای از مشاهدات است که بر حسب زمان و یا هر کمیت دیگری مرتب شده باشند. اساس بسیاری از تصمیم‌گیری‌ها در فرایندهای هیدرولوژیکی و تصمیمات بهره‌برداری از منابع آب بر پایه پیش‌بینی و تحلیل سری‌های زمانی می‌باشد (Karamouz و Araghinajad، ۲۰۰۵). کاربرد سری‌های زمانی در هیدرولوژی از چهار دهه پیش آغاز شده و با ارائه مدل‌های باکس و جنکینز به اوج خود رسید (Azizi، ۲۰۰۵). پژوهش‌های گسترده‌ای در دنیا در زمینه پیش‌بینی جریان انجام شده که نشانگر اهمیت موضوع پیش‌بینی جریان ورودی به سدها می‌باشد. Kamali (۲۰۰۲)، با استفاده از سری‌های زمانی اقدام به پیش‌بینی جریان ماهانه ورودی به مخزن سد جیرفت کرد. با توجه به نتایج به‌دست آمده مدل ARIMA(1,1,1)(0,1,2) (مدل میانگین متحرک خودهمبسته یکپارچه[1]) به‌عنوان مدل برتر انتخاب شد و در نهایت با استفاده از این مدل دبی ماهانه دو سال آینده پیش‌بینی شد. Yourekli و همکاران (۲۰۰۴) کارایی مدل‌های ARIMA و مدل‌های توماس-فرینگرا در پیش‌بینی مقادیر بیشینه ماهانه در سه زیرحوضه رودخانه cekerek در شمال آناتولی ترکیه بررسی کردند. نتایج آن‌ها نشان داد، مدل‌های سری‌های زمانی ARIMA عملکرد بهتری نسبت به مدل‌های توماس و فرینگ داشته است. Hsanah و همکاران (۲۰۱۳) برای پیش‌بینی سیلاب از مدل تابع انتقال استفاده کردند و به این نتیجه رسیدند که مدل تابع انتقال نسبت به مدل‌های سری زمانی آریما پیش‌گویی بهتری دارد. Valipour و همکاران (۲۰۱۳) از مدل‌های آرما، آریما و مدل شبکه عصبی مصنوعی[2] خود همبسته، برای مدل‌سازی جریان ماهانه ورودی به مخزن سد دز استفاده کردند. طبق نتایج حاصله، افزایش مرتبه مدل‌های سری زمانی باعث بهبود دقت مدل‌سازی شده و مدل آریما نسبت به آرما با دقت بیشتری توانسته جریان ماهانه را پیش‌بینی نماید. به‌طور کلی، مدل شبکه عصبی مصنوعی خودهمبسته بهتر از مدل‌های سری زمانی جریان ورودی ماهانه به سد دز را مدل‌سازی کرده است. Estaki و همکاران (۲۰۱۶)، اقدام به پیش‌بینی دبی ایستگاه پل فسا با استفاده از سری‌های زمانی کردند. نتایج آن‌ها نشان داد که مدل MA(1,1,1) بهترین مدل برای پیش‌بینی دبی ماهانه در این ایستگاه می‌باشد. در سال‌های اخیر استفاده از روش‌های هوش مصنوعی از جمله منطق فازی و شبکه عصبی در زمینه‌های مختلفی که در آن‌ها ارتباط بین ورودی و خروجی غیر خطی بوده، گسترش پیدا کرده است. در پژوهش حاضر، جهت مقایسه نتایج مدل‌سازی با شبکه عصبی مصنوعی و همچنین، سری‌های زمانی، سعی شد که عملکرد هر دو روش در مدل‌سازی و پیش‌بینی جریان رودخانه با استفاده از داده‌های یکسان مورد بررسی قرار گیرد. همچنین، در این پژوهش، از مدل ماشین پشتیبان تصمیم[3] (SVM) به‌منظور پیش‌بینی دبی استفاده شد. مدل‌های SVM که بر پایه تئوری آموزش احتمال قرار دارند، یک طبقه جدید از مدل‌هایی هستند که به‌منظور طبقه‌بندی و پیش‌بینی در علوم مختلف استفاده می‌شوند. Cheng و همکاران (۲۰۱۵)، با استفاده از سه مدل شبکه عصبی مصنوعی، ماشین بردار پشتیبان و مدل ترکیبی اقدام به پیش‌بینی دبی ماهانه سد خینفنجیانگ[4] در چین کردند. نتایج آن‌ها نشان داد که هر سه مدل نتایج

[1] Autoregressive integrated moving average

[2] Artificial neural network
[3] Support Vector Mashie (SVM)
[4] Xinfengjiang

www.SID.ir

Archive of SID

رضایت‌بخشی داشته‌اند. همچنین، بر اساس شاخص‌های ارزیابی در حالتی که مدل‌ها به‌صورت ترکیبی به‌کار رفته‌اند، نتایج بهتری دارد، نسبت به حالتی که مدل‌ها به‌صورت تکی به‌کار می‌روند. Ghorbani و همکاران (۲۰۱۶)، اقدام به ارزیابی روش‌های هیدرولوژیکی و داده‌کاوی در شبیه‌سازی و پیش‌بینی دبی جریان ماهانه در ایستگاه هیدرومتری ارازکوسه کردند. نتایج آن‌ها نشان داد، کاربرد مدل‌های داده‌کاوی منجر به بهبود نتایج شبیه‌سازی می‌شود. Young و همکاران (۲۰۱۷) از یک مدل ترکیبی برای پیش‌بینی دبی در حوضه Chishan Creek در جنوب تایوان استفاده کردند. نتایج آن‌ها نشان داد که مدل ترکیبی نسبت به سه مدل فردی هیدرولوژیکی، شبکه عصبی مصنوعی و ماشین بردار پشتیبان نتایج بهتری دارد. Yu و همکاران (۲۰۱۸) از مدل داده‌کاوی رگرسیون بردار پشتیبان برای پیش‌بینی دبی ماهانه ورودی به سه سد Gorges در چین استفاده کردند. نتایج آن‌ها نشان داد که مدل مورد استفاده به‌خوبی قادر است، دبی ماهانه ورودی به سد را پیش‌بینی کند. یکی از ابزارهای اساسی در مدیریت منابع آب پیش‌بینی عرضه و تقاضای آب می‌باشد. لزوم پیش‌بینی مناسب جریان رودخانه در کارهای عمرانی، ساماندهی رودخانه، سیستم‌های هشدار سیل و به‌خصوص برنامه‌ریزی برای بهره‌برداری بهینه از مخازن سدها کاملا احساس می‌شود. در این راستا، استفاده از روش‌هایی که بتواند با کمینه خطا و با توجه به داده و اطلاعات موجود، جریان رودخانه را پیش‌بینی کند، از اهمیت فراوانی برخوردار می‌باشد. در این میان، روش‌های سری‌های زمانی و شبکه عصبی مصنوعی از جمله روش‌های متداول در پیش‌بینی دبی رودخانه بوده و به‌عنوان یک روش استاندارد در بین هیدرولوژیست‌ها شناخته شده است. سد بوستان که بر روی گرگان‌رود واقع شده، به‌دلیل کنترل سیلاب‌ها و کاهش خسارات ناشی از آن‌ها و همچنین، توسعه کشت آبی اراضی ساحل راست گرگان‌رود دارای اهمیت زیادی می‌باشد. بنابراین، هدف از انجام این پژوهش بررسی کارایی مدل‌های سری زمانی، شبکه عصبی مصنوعی، SVM و مدل‌های ترکیبی در پیش‌بینی دبی ورودی به سد بوستان در استان گلستان و در نهایت مقایسه نتایج مدل‌های مختلف و ارائه روش مناسب و با دقت کافی برای پیش‌بینی جریان ورودی به مخزن سد بوستان می‌باشد.

مواد و روش‌ها

منطقه مورد پژوهش: سد مخزنی بوستان در موقیت جغرافیایی ″۲۹ ′۲۵ °۵۵ طول شرقی و ″۳۰ ′۲۵ °۳۷ عرض شمالی قرار گرفته است. این سد در ۳۵ کیلومتری شمال شرقی گنبد کاووس بر روی گرگان‌رود و در داخل حوضه گرگان‌رود واقع شده است. سد مذکور از نوع خاکی همگن بوده و دارای بیشینه ارتفاع ۳۵ متر از پی و ۲۵ متر از کف مخزن می‌باشد. طول تاج سد ۶۶۵ متر و عرض تاج آن ۱۰ متر است. ایستگاه تمر که در خروجی حوزه آبخیز سد بوستان و بر روی رودخانه گرگان‌رود واقع شده است، به‌عنوان تنها ایستگاه شاخص برای اندازه‌گیری میزان آب ورودی به مخزن سد بوستان می‌باشد. این ایستگاه در موقعیت جغرافیایی ″۲۱ ′۳۰ °۵۵ طول شرقی و ″۰۵ ′۲۹ °۳۷ عرض شمالی قرار گرفته و ارتفاع آن از سطح دریا ۱۱۷ متر می‌باشد. شکل ۱، موقعیت کلی حوضه، سد و ایستگاه مورد مطالعه را نشان می‌دهد.

پیش‌بینی دبی با استفاده از سری‌های زمانی: برای پیش‌بینی دبی از سری داده‌های دبی ماهانه ایستگاه تمر با دوره آماری ۴۷ سال (۱۳۴۸ تا ۱۳۹۴) استفاده شد. پیش‌بینی با مدل سری‌های زمانی به‌صورت زیر انجام شد. ابتدا، نرمال بودن سری داده‌ها بررسی شد. بدین منظور، ضریب چولگی داده‌ها را محاسبه کرده و با استفاده از رابطه (۱) نرمال بودن داده‌ها تعیین شد.

$$-\mu_{1-\alpha/2}\sqrt{\frac{6}{n}} < \chi < \mu_{1-\alpha/2}\sqrt{\frac{6}{n}} \qquad (۱)$$

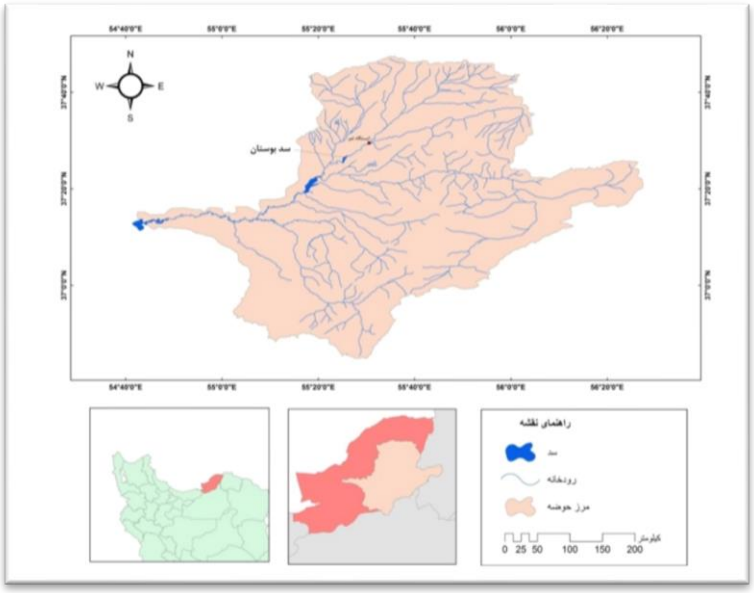
که در آن، χ ضریب چولگی داده‌ها می‌باشد. اگر داده‌ها در این مرحله نرمال نباشند، با استفاده از یک تابع مناسب داده‌ها نرمال می‌شود. بدین منظور، از جمله تابع‌های متداول برای نرمال‌سازی می‌توان به تابع‌های لگاریتمی، تبدیل کاکس و باکس، جذری و غیره اشاره کرد. در این پژوهش، با توجه به اینکه داده‌ها نرمال نبودند از تبدیل باکس و کاکس (رابطه ۲) برای نرمال‌سازی استفاده شد.

www.SID.ir

Archive of SID

$$y=\begin{cases}=\lambda^{-1}\left[(x+c)^{\lambda}-1\right] & \lambda\neq 0\\ =\ln(x+c) & \lambda=0\end{cases} \qquad (۲)$$

مرحله بعد محاسبه و رسم کلوروگرام خودهمبستگی و خودهمبستگی جزئی می‌باشد که از ترسیم مقادیر خودهمبستگی و خودهمبستگی جزئی در تأخیرهای مختلف به‌دست می‌آید. شکل کلوروگرام تاثیر به‌سزایی در انتخاب نوع مدل و مرتبه مدل دارد.



شکل ۱- موقعیت کلی حوضه گرگان‌رود، موقعیت سد و ایستگاه مورد مطالعه

بعد از تعیین نوع مدل و مرتبه آن، مقادیر پارامترهای مدل را تخمین زده که برای این منظور معمولا از دو روش مومنتم و بیشینه درست‌نمایی استفاده می‌شود. معمولا از روش مومنتم برای تخمین اولیه پارامترها استفاده شده و مقادیر نهایی پارامترها با استفاده از روش بیشینه درست‌نمایی تعیین می‌شود. در این پژوهش، برای تعیین پارامترها از نرم افزار MINITAB استفاده شد که در این نرم‌افزار از روش بیشینه درست‌نمایی برای تعیین پارامترها استفاده می‌شود. برای آزمایش درستی مدل محاسبه شده، ابتدا باقی‌مانده‌ها با استفاده از رابطه (۳) محاسبه شد.

$$E_k=Z_t-Z_t(k) \qquad (۳)$$

که در آن، E_K خطای kام می‌باشد که از تفریق Z_t از مقدار پیش‌بینی شده آن به‌دست می‌آید و در آن فرض می‌شود، باقی‌مانده‌ها یک نویز سفید با توزیع نرمال می‌باشد. پس از آنکه مقادیر باقی‌مانده‌ها محاسبه شد، برای بررسی استقلال باقی‌مانده‌ها از توابع خودهمبستگی و خودهمبستگی جزئی و همچنین، آماره باکس پیرسون ۴ استفاده می‌شود. با مقایسه Q با مقدار به‌دست آمده از جدول کای دو با درجه آزادی PL- (که L تعداد تاخیرها و p تعداد پارامترها می‌باشند) می‌توان پی به استقلال باقی‌مانده‌ها برد. در صورتی‌که $Q<\chi^2(l-p)$ باشد، غیر وابسته بودن مقادیر باقی‌مانده‌ها ثابت شده و مدل پذیرفته می‌شود. در صورتی‌که خلاف رابطه برقرار شده باشد، مدل دیگر با رتبه بالاتر (p+1) را می‌بایست بررسی کرد.

$$Q=n(n+2)\sum_{k=1}^{k}\frac{r_k^2}{n-k} \qquad (۴)$$

معمولا برای هر سری زمانی چند مدل مختلف وجود دارد که صحت آن‌ها نیز تعیین شده در نتیجه مدلی به‌عنوان مدل نهایی انتخاب می‌شود که ممسک‌تر[1] باشد که برای این منظور از معیار آکائیک[2] (رابطه ۵) استفاده می‌شود.

$$AIC(M)=n\ln\sigma_a^2+2M \qquad (۵)$$

که در آن‌ها، مقدار σ_a^2 واریانس تخمینی خطا با

[1] Parsimony
[2] Akaike information criterion

www.SID.ir

Archive of SID

استفاده از روش بیشینه درست‌نمایی، n تعداد داده‌ها و M تعداد پارامترهای مدل می‌باشد. از آن‌جایی که معیار آکائیک مرتبه مدل‌های اتورگرسیو را بالا نشان می‌دهد، بنابراین، در این پژوهش برای انتخاب مدل ممسک‌تر از معیار شوارتز (رابطه ۶) نیز استفاده شد.

$$SBC(M) = n\ln\sigma_a^2 + M\ln n \qquad (۶)$$

در این مرحله، هر یک از مدل‌ها که مقدار اکائیک و شوارتز آن‌ها پایین‌تر باشد، به‌عنوان مدل برتر برای مراحل بعدی انتخاب شد. در نهایت با استفاده از مدلی که درستی آن تعیین و با توجه به معیارهای آکائیک و شوارتز به‌عنوان مدل ممسک‌تر انتخاب شد، اقدام به پیش‌بینی دبی و تولید داده‌های مصنوعی دبی شد. از آنجایی که مدل‌های سری‌های زمانی در بیشتر از یک گام زمانی دارای پیش‌بینی مناسبی نبوده، فقط مقادیر دبی در یک گام زمانی آینده پیش‌بینی شده است. پس از هر مرحله از پیش‌بینی، مقادیر پارامترهای مدل را دوباره تعیین کرده و به اصطلاح پیش‌بینی‌ها به هنگام می‌شوند. در پژوهش حاضر، در مراحل مختلف برای پیش‌بینی با سری زمانی از نرم‌افزار MINITAB استفاده شد.

پیش‌بینی دبی با استفاده از شبکه عصبی مصنوعی: در روش‌های هوش مصنوعی دو مرحله کلی پیشپردازش و تعیین ساختار بهینه شبکه وجود دارد. پیش پردازش را می‌توان به‌عنوان هر گونه عملیات بر روی داده‌های خام که باعث افزایش کارایی شبکه در مدل‌سازی داده‌های ورودی و خروجی می‌شود، تفسیر کرد. در صورتی‌که داده‌ها به‌صورت خام وارد شبکه شوند، به‌علت تغییرات زیاد داده‌ها تاثیر متفاوتی بر روی شبکه گذاشته به‌طوری که برخی از نورون‌ها خیلی زود به حد آتش رسیده، در حالی‌که برخی دیگر از نورون‌ها حتی به آستانه فعالیت نیز نرسیده‌اند. لذا، در ابتدا لازم است که داده‌ها معیارسازی شوند، یعنی بین یک دامنه عددی که معمولا صفر و یک می‌باشد، قرار گیرند. اگر تابع سیگموئید را در نظر بگیریم، مشاهده می‌شود، شیب این تابع در محدوده صفر تا یک زیاد بوده و به‌تدریج که به صفر و یک نزدیک می‌شود، شیب آن کاهش پیدا می‌کند. در این ارتباط و با توجه به تحقیقات قبلی (Eslami، ۲۰۰۲؛ Khisi، ۲۰۰۴؛ Salamatian، ۲۰۰۶) به‌خاطر این‌که مقادیر آستانه صفر و یک را در داده‌های ورودی به شبکه نداشته باشیم، داده‌ها با استفاده از رابطه (۷) که باعث می‌شود، داده‌ها بین ۰/۰۵ و ۰/۹۵ قرار گیرند، معیارسازی شده‌اند.

$$y = 0.9 \times \frac{X_i - X_{\min}}{X_{\max} - X_{\min}} + 0.05 \qquad (۷)$$

که در آن، $X_{\min}$ کوچک‌ترین داده و $X_{\max}$ بزرگ‌ترین داده سری داده‌ها می‌باشد. با استفاده از این رابطه داده‌های ورودی بین ۰/۰۵ و ۰/۹۵ قرار می‌گیرند. پس از معیارسازی باید الگوهای آموزشی، صحت‌سنجی و تست انتخاب شود. از الگوهای آموزش ده برای آموزش شبکه استفاده می‌شود که شبکه با استفاده از الگوهای آموزشی وزن و بایاس خود را تنظیم می‌کند. بنابراین، بیشترین درصد از داده‌ها مربوط به الگوهای آموزشی می‌باشد. از آنجا که شبکه عصبی فاقد برون‌یابی بالاست و توانایی تعمیم‌دهی آن فقط در چارچوب درون‌یابی مطرح است، بنابراین، داده‌های آموزشی طوری باید انتخاب شوند که نماینده کل داده‌ها باشند. به‌عبارت دیگر، تمام حالات ممکن در داده‌ها از قبیل مقادیر حدی (بیشینه و کمینه) را در خود داشته باشند. از آنجایی که در این پژوهش هدف مقایسه عملکرد روش‌های هوش مصنوعی با مدل‌های سری‌های زمانی می‌باشد و در مدل‌های سری‌های زمانی ترتیب داده‌ها برای ما مهم می‌باشد، لذا، در این پژوهش از رویکرد تصادفی نمی‌شود برای تعیین ورودی‌ها استفاده کرد. بدین منظور، تا حد ممکن باید درصد داده‌های آموزشی را بالا گرفت (۷۰ درصد از داده‌های کل). داده‌های تصدیق در هنگام آموزش از افزایش خطای مدل به‌دلیل هم‌گرایی بیش از حد به داده‌های آموزش، جلوگیری می‌کند. در حالی که، درصد خطای نسبی داده‌های آموزش کاهش می‌یابد، درصد خطای نسبی داده‌های تصدیق افزایش پیدا می‌کند. در این مواقع مدل با داده‌های آموزش بیشتر منطبق می‌شود، لذا لازم است که از آموزش بیشتر شبکه جلوگیری شود. در این حالت، بهترین نقطه (آموزش) نقطه کمینه خطای نسبی مربوط به داده‌های تصدیق می‌باشد. در این پژوهش، ۱۵ درصد کل داده‌ها را برای تصدیق در نظر گرفته شده که البته این ۱۵ درصد داده‌های تصدیق در روش‌های

www.SID.ir

Archive of SID

سری‌های زمانی به‌عنوان قسمتی از داده‌های آموزش می‌باشد. داده‌های تست را می‌توان به‌عنوان نوعی پس‌پردازش در نظر گرفت که با استفاده از این داده‌ها کارایی شبکه‌های مختلف بررسی می‌شود. در این پژوهش، ۱۵ درصد آخر داده‌ها به‌عنوان داده‌های تست در نظر گرفته می‌شود. در شبکه عصبی ساختار شبکه شامل نوع ورودی‌ها و خروجی‌ها، تعداد لایه‌های مخفی و تعداد نورون‌ها در هر لایه مخفی، نوع تابع فعال‌سازی و نوع الگوریتم آموزشی می‌باشد. لازم است تا روابط ما بین ورودی‌ها و خروجی‌ها تعیین و بهترین ورودی‌ها تعیین شود. برای تعیین بهترین ورودی دو نظریه وجود دارد. در یک روش، استفاده از برخی پیش پردازش‌های آماری از قبیل همبستگی عرضی، خودهمبستگی و خودهمبستگی جزئی بین داده‌ها پیشنهاد شده است (Nanduri و Swain، ۲۰۰۵؛ El-Shafie و همکاران، ۲۰۰۷) و در گروهی دیگر اعتقاد دارند، با توجه با روابط غیر خطی بین ورودی‌ها و خروجی‌ها و خطی بودن روش‌هایی از قبیل همبستگی عرضی، خودهمبستگی و خودهمبستگی جزئی بهترین ورودی‌ها باید به‌وسیله سعی و خطا تعیین شوند (Khisi، ۲۰۰۴؛ Jain و Kumar، ۲۰۰۷؛ Firat و Gungor، ۲۰۰۷). در این پژوهش، برای روش‌های هوش مصنوعی، ورودی‌های شبکه مقادیر گذشته دبی بود. در ابتدا ورودی شبکه را مقادیر دبی تا تاخیر پنج در نظر گرفته شد که به‌صورت جدول ۱ می‌باشد.

جدول ۱- ترکیب ورودی‌های مختلف در مدل شبکه عصبی مصنوعی

نوع ورودی	نام مدل
Q_{r-1}	ورودی ۱
Q_{t-2} و Q_{r-1}	ورودی ۲
Q_{t-3} و Q_{t-2} , Q_{r-1}	ورودی ۳
Q_{t-4} و Q_{t-3} , Q_{t-2} , Q_{r-1}	ورودی ۴
Q_{t-5} و Q_{t-4} , Q_{t-3} , Q_{t-2} , Q_{r-1}	ورودی ۵

در شبکه عصبی پرسپترون چند لایه، یک لایه ورودی و یک لایه خروجی و یک یا چند لایه مخفی وجود دارد که تعداد نرون‌ها در لایه‌های ورودی و خروجی برابر با تعداد بردارهای ورودی و خروجی می‌باشد. مسئله اصلی در این شبکه‌ها، تعیین تعداد لایه‌های پنهان و تعداد نورون‌های آن‌ها می‌باشد که در این رابطه نظرات متفاوتی وجود دارد. هر چه تعداد لایه‌های مخفی بیشتر باشد، قدرت تجزیه و تحلیل شبکه بیشتر می‌شود، ولی در صورت زیاد شدن تعداد لایه‌های میانی سرعت شبکه برای هم‌گرایی به جواب، بسیار پایین می‌آید. در ارتباط با نوع تابع فعال‌سازی با توجه به پژوهش‌های قبلی نوع تابع فعال‌سازی سیگموئید استفاده شد (Rezaei، ۲۰۰۵). نوع الگوریتم آموزش نیز با توجه به کارایی روش پس انتشار خطا در مطالعات هیدرولوژیکی، از این الگوریتم آموزش استفاده شد. در این پژوهش، به‌دلیل جامعیت و سهولت کاربردی از نرم‌افزار MATLAB استفاده شد. در این نرم‌افزار، برای ساخت مدل‌های مختلف شبکه عصبی از جعبه ابزار Artificial Neural Network استفاده شد.

پیش‌بینی دبی با استفاده **از مدل SVM:** یکی از روش‌های کارآمد جهت مدل‌سازی داده پیوسته و گسسته مبتنی بر آموزش آماری روش ماشین بردار پشتیبان است. در این روش، داده‌هایی که در بعد پایین غیر خطی هستند، با استفاده از تابع کرنل به بعد بالاتر انتقال و در این بعد قابل مدل‌سازی به‌صورت خطی می‌باشند (Dibike و همکاران، ۲۰۰۱). در مدل بردار پشتیبان رگرسیونی فرض می‌شود که تعداد N سری داده آموزشی $\{(x_i, d_i)\}_i^N$ موجود باشد. معادله رگرسیون خطی که در فضای ویژگی با بعد بالا به داده‌ها برازش داده می‌شود، به‌صورت رابطه (۸) است.

$$y_i = f(x) = \langle w_i . \emptyset_i(x) \rangle + b \qquad (۸)$$

www.SID.ir

Archive of SID

که در آن، w_i و b ضرایب رابطه در فضای ویژگی در بعد بالا و $\emptyset_i(x)$ تابعی است که داده‌ها را از فضای ورودی به فضای ویژگی افراز می‌کند. مانند هر مدل دیگر برای تخمین پارامترهای این مدل باید یک معیار خطا در نظر گرفت. در این رابطه، توابع خطای مختلف مثل خطی، کوادراتیک، توانی و غیره استفاده شده است. در این تحقیق، از معیار Vapnik's-ε که به‌صورت رابطه (۹) ارائه شده است، برای بهینه‌سازی پارامترهای مدل بالا استفاده شد.

$$L_\varepsilon(v, g(u)) = \begin{cases} 0 & for \ \ |v - g(u)| \le \varepsilon \\ |d - g(u)| - \varepsilon & otherwise \end{cases} \quad (۹)$$

با در نظر گرفتن این معیار خطا، تخمین پارامترهای رابطه رگرسیونی به مسئله بهینه‌سازی Convex تبدیل که به صورت روابط (۱۰و ۱۱) بیان می‌شود.

$$min \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n}(\xi_i + \xi_i^*) \quad (۱۰)$$

$$subject\ to \begin{cases} v_i - (w.\emptyset(u_i) + b) \le \varepsilon + \xi_i \\ (w.\emptyset(u_i) + b) - v_i \le \varepsilon + \xi_i^* \\ \xi_i, \xi_i^* \ge 0, \quad i = 1,2,\dots.,n \end{cases} \quad (۱۱)$$

که در آن‌ها، ξ_i^* و ξ_i ضرایب کمبود[1] (که حد بالا و پایین خطای آموزش را تعیین می‌کند)، C پارامتر تنظیم‌کننده[2] و ε حد بالای خطای خروجی هستند. با در نظر گرفتن ضرایب لاگرانژ و شرایط خان تاکر فرم دوگانه[3] مسئله بهینه‌سازی بالا تبدیل به مسئله بهینه‌سازی برنامه‌ریزی کوادراتیک می‌شود که با بهینه‌سازی آن ضرایب لاگرانژ $\underline{a}_i$، $\overline{a}_i$ و بردار وزن‌های بهینه ابر صفحه رگرسیون محاسبه و در نهایت برای بردارهای پشتیبان رابطه رگرسیون به‌صورت رابطه (۱۲) بیان می‌شود.

$$g(x) = \sum_{i=1}^{nl}(\underline{a}_i - \overline{a}_i)K\langle\Phi(x_i), \Phi(x_j)\rangle + b \quad (۱۲)$$

که در آن، nl نشان‌دهنده تعداد بردار پشتبان است. در رابطه‌ها ضرب داخلی در فضای ویژگی به‌وسیله تابع کرنل صورت می‌گیرد، یعنی چنان‌چه دو بردار $\Phi(x_i)$ و $\Phi(x_j)$ وجود داشته باشد، ضرب داخلی آن‌ها به‌وسیله تابع کرنل به‌صورت رابطه صورت می‌گیرد. هر تابعی که در شرایط Mercer صدق کند، می‌تواند به‌عنوان تابع کرنل مورد استفاده قرار گیرد (Vapnik، ۱۹۹۸). از توابع کرنل متداول می‌توان به تابع‌های کرنل خطی، چند جمله‌ای، گوسی (رابطه ۱۳) نام برد که در این پژوهش از تابع گوسی استفاده شد.

$$K(x_i, x_j) = exp(-\|x_i - x_j\|/2\sigma^2) \quad (۱۳)$$

در روش ماشین بردار پشتیبان پارامترهای مدل شامل مقادیر C، e و پارامترهای تابع کرنل آن می‌باشد. جهت تعیین این پارامترهای الگوریتم مشخصی وجود ندارد، بدین منظور از روش سعی و خطا و به‌صورت جستجوی شبکه‌ای مقادیر پارامترهای آن تعیین شد. در روش جستجوی شبکه‌ای به‌دلیل اینکه هیچ الگوریتم بهینه‌سازی جهت‌دار برای بهنیه‌سازی پارامترها استفاده نمی‌شود، نیاز است تا مدل به تعداد خیلی زیاد اجرا شود. برای رفع این مشکل از الگوریتم جستجوی شبکه‌ای دو مرحله‌ای استفاده شد. در این الگوریتم، در اولین گام جستجوی مقادیر پارامترها در دامنه بزرگ‌تر، اما با دقت کمتری انجام می‌گیرد. در دومین مرحله جستجوی به‌صورت جزئی‌تر و در قسمتی از فضای پارامترها که در مرحله قبل بهتر جواب داده صورت می‌گیرد (Chen و Yu، ۲۰۱۰). برای پیش‌بینی با مدل SVM از نرم‌افزار R استفاده شد. در این مدل، همانند شبکه عصبی مصنوعی پنج ورودی از ورودی‌های یک تا پنج استفاده شد. بدین صورت که با استفاده از هر پنج ورودی اقدام به پیش‌بینی شد و سپس با استفاده از معیارهای ارزیابی بهترین مدل برای هر ایستگاه تعیین شد و در نهایت با استفاده از مدل برتر نهایی هیدروگراف و دیاگرام پراکنش برای ایستگاه تمر ترسیم شد.

روش های ترکیبی: در پیش‌بینی دبی انتخاب یک مدل مناسب کار مشکلی می‌باشد، زیرا از یک‌سو، مدلی که نسبت به سایر مدل‌ها در تمام حوضه‌ها و شرایط، برتری مطلق داشته باشد، وجود نداشته و از سوی دیگر، تمام مدل‌ها یک مقدار عدم قطعیت شامل عدم قطعیت در پارامترها و ساختار دارند. در نتیجه به جای استفاده از یک روش منفرد می‌توان مدل‌ها مختلف را ترکیب کرد، به‌طوری که یک پیش‌بینی مناسب‌تر و قابل اعتمادتری نسبت به هر یک از مدل‌های منفرد به‌دست آید. در روش‌های ترکیبی

[1] Slack variables
[2] Regularization parameter
[3] Dual

www.SID.ir

Archive of SID

فرض شده که هر سری زمانی از دو قسمت خطی و غیر خطی تشکیل شده است. به‌خاطر تغییرات زیاد در الگوهای داده‌های سری زمانی استفاده از روش‌های هیبرید می‌تواند عدم قطعیت مدل را که به‌طور نمونه در پارامترهای آماری سری‌های زمانی اتفاق می‌افتد را کاهش دهد. در این پژوهش، از سه روش ترکیبی استفاده شده است که به‌ترتیب به شرح زیر می‌باشد.

روش ترکیبی ۱: این روش اولین‌بار به‌وسیله Zhang (۲۰۰۳) پیشنهاد شد. در این روش، فرض شده که هر سری زمانی از دو قسمت خطی و غیر خطی تشکیل شده است (رابطه ۱۴).

$$Y_t = L_t + N_t \qquad (۱۴)$$

که در آن، L_t جزء خطی و N_t جزء غیر خطی می‌باشد. در این روش، ابتدا قسمت خطی به‌وسیله مدل ARIMA مدل‌سازی شده و سپس، باقی‌مانده‌های مرحله قبل (که حاوی روابط غیر خطی بوده) (رابطه ۱۵) به‌وسیله شبکه عصبی مدل‌سازی می‌شود.

$$e = Y_t - \hat{L}_t \qquad (۱۵)$$

که در آن، $\hat{L}_t$ به‌وسیله مدل ARIMA تخمین زده می‌شود. در مدل‌های خطی برای این‌که یک مدل برازش داده شده به داده‌ها درست باشد، باید هیچ‌گونه همبستگی خطی در باقی‌مانده‌ها موجود نباشد. با مدل کردن باقی‌مانده‌ها با استفاده از روش‌های هوش مصنوعی روابط غیر خطی موجود در باقی‌مانده‌ها را می‌توان تعیین کرد.

روش ترکیبی ۲: در این روش، خروجی نهایی حاصل خروجی‌های وزن‌دار شده هر یک از مدل‌های منفرد می‌باشد (رابطه ۱۵).

$$y_{t+1} = \sum_{j=1}^{k} a_j f_{t,j} + \varepsilon_{t+1} \qquad (۱۶)$$

که در آن، $f_{t,j}$ مقادیر پیش‌بینی شده در یک گام زمانی آینده از y_{t+1} به‌وسیله روش iام در زمان t، a_j مقادیر وزن‌ها و ε_{t+1} مقدار باقی‌مانده می‌باشد.

روش ترکیبی ۳: در این روش، مقادیر خروجی هر سه روش به‌عنوان ورودی یکی از روش‌های هوش مصنوعی در نظر گرفته می‌شود و با استفاده از معیارهای ارزیابی بهترین روش هوش مصنوعی و ساختار مربوط به آن تعیین می‌شود. در این پژوهش، مقادیر خروجی هر سه روش به‌عنوان ورودی وارد مدل SVM شد و مدل‌سازی انجام شد.

معیارهای ارزیابی: در هر مرحله برای تعیین بهترین ساختار مدل و مقایسه مدل‌های مختلف از معیارهای زیر استفاده شده است.

- ضریب همبستگی
- معیار RMSE (میانگین مربعات خطا[1])

$$RMSE = \frac{1}{n}\sqrt{\sum_{i=1}^{n}(Q_t^o - Q_t^c)^2} \qquad (۱۷)$$

- معیار MAE (میانگین مطلق خطا[2])

$$MAE = \frac{1}{n}\sum_{1}^{n}\left[Q_t^o - Q_t^c\right] \qquad (۱۸)$$

- معیار AARE[3] (متوسط مقادیر مطلق خطای نسبی)

$$AARE = \frac{1}{N}\sum_{1=1}^{n}\left|\frac{(Q_t^o - Q_c)}{Q_t^o}\times 100\right| \qquad (۱۹)$$

- معیار MBE (میانگین انحراف خطا[4])

$$MBE = \frac{1}{n}\sum_{i=1}^{n}(\mathrm{Qobs} - \mathrm{Qsim}) \qquad (۲۰)$$

- معیار CE (معیار نش-ساتکلیف[5])

$$\mathrm{CE} = 1 - \frac{\Sigma(\mathrm{Q_{obs}} - \mathrm{Q_{sim}})^2}{\Sigma(\mathrm{Q_{obs}} - \bar{\mathrm{Q}})^2} \qquad (۲۱)$$

- معیار خطای PE (خطای پیک[6])

$$PE = \frac{\left|(Q_p)_{obs} - (Q_p)_{sim}\right|}{(Q_p)_{obs}} \qquad (۲۲)$$

که در آن‌ها، n تعداد داده ها، Q_t^o دبی مشاهداتی و Q_t^c دبی شبیه‌سازی شده می‌باشند.

نتایج و بحث

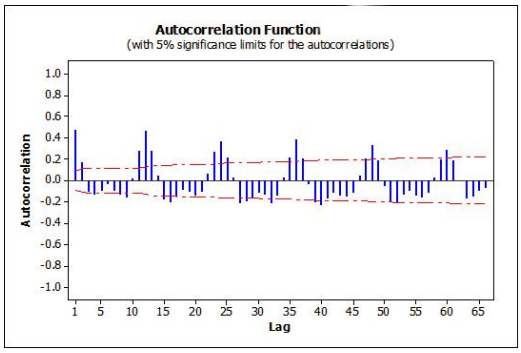
برای سری‌های زمانی ابتدا نرمال بودن داده‌ها تعیین شد که با توجه به ضریب چولگی بالای داده‌ها در ایستگاه مورد مطالعه داده‌ها نرمال نبودند و برای نرمال‌سازی، از روش باکس-کاکس استفاده شد. در مرحله بعد، مقادیر ضرایب خودهمبستگی و خود-همبستگی جزئی در تاخیرهای مختلف تعیین و

[1] Root Mean Square Error
[2] Mean Absolute Error
[3] Average Absolute Relative Error
[4] Mean Bias Error
[5] Nash–Sutcliffe Efficiency Coefficient
[6] Peak Error

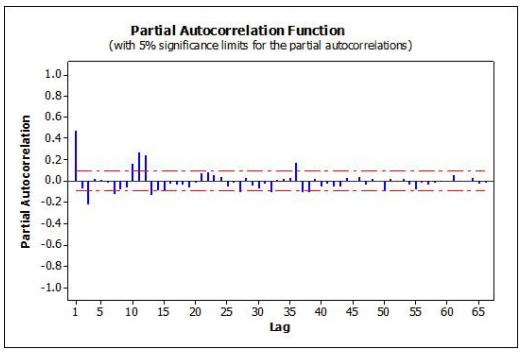
www.SID.ir

Archive of SID

محاسبه شد (شکل‌های ۲ و ۳). با توجه به ضریب خودهمبستگی ملاحظه می‌شود که نرخ کاهش مقادیر ضریب خودهمبستگی در تاخیرهای مختلف بسیار پایین بوده و به‌صورت سینوسی نوسان داشته که این امر حاکی از غیر ایستا بودن سری زمانی مورد مطالعه دارد. همچنین، ضرایب خودهمبستگی در تاخیرهای مختلف نوسان داشته و در تاخیر مشخصی بریده نشده است. این امر نشان‌دهنده وجود مدل‌های اتورگرسیو در سری مورد مطالعه می‌باشد. ضرایب خودهمبستگی جزئی در تاخیرهای ۱ بریده شده و همچنین، مقادیر ضرایب خودهمبستگی جزئی در تاخیرهای ۱۲ بزرگ می‌باشد. بنابراین، در گام اول مدل‌های ARIMA به‌عنوان مدل مناسب تشخیص داده شده و در مرحله بعد مدل‌ها مورد بررسی بیشتر قرار گرفته و با استفاده از بیشینه درست‌نمایی، پارامترهای هر یک از مدل‌ها تعیین شد.



شکل ۲- خود همبستگی داده‌های ایستگاه تمر



شکل ۳- خود همبستگی جزئی داده‌های ایستگاه تمر

نتایج مدل‌های ARIMA در جدول ۲ آورده شده است. برای تعیین صحت هر یک از مدل‌های تعیین شده، از آزمون‌های باکس-پیرسون، ضرایب خودهمبستگی و خودهمبستگی جزئی باقی‌مانده‌ها استفاده شد. با توجه به جداول، بر اساس معیارهای آکاییک و شوارتز در ایستگاه تمر مدل (1,0,1) ARIMA (2,0,0) به‌عنوان مدل ممسک‌تر انتخاب شد. پس از این‌که مدل مناسب انتخاب شد، اقدام به پیش-بینی مقادیر آینده دبی شد. با توجه به این‌که مدل-های سری زمانی در پیش‌بینی‌های بیشتر از یک گام زمانی بسیار بد عمل می‌کنند، بنابراین، فقط یک گام زمانی آینده پیش‌بینی شد و اصطلاحاً در هر گام زمانی پیش‌بینی به هنگام شد. به این صورت که در هر گام زمانی برای پیش‌بینی مقدار دبی آینده در ابتدا با استفاده از مقادیر قبلی دبی پارامترهای مدل منتخب تعیین و سپس یک گام زمانی آینده پیش‌بینی شد. در مرحله بعد، برای پیش‌بینی یک گام زمانی بعد مقدار مشاهداتی دبی در گام زمانی قبل را به داده‌ها اضافه کرده و پارامترهای مدل برای یک گام زمانی آینده تعیین شد. در شبکه عصبی مصنوعی ساختار شبکه شامل تعداد لایه‌های مخفی، تعداد نرون‌های هر یک از لایه‌های مخفی، نوع تابع فعال‌سازی و ورودی‌های مدل که بر روی خروجی مدل تاثیر گذاشته، باید تعیین شود. برای این منظور، در ایستگاه مورد نظر شبکه‌هایی با یک لایه مخفی بررسی شد. با توجه به نتایج به‌دست آمده در ایستگاه تمر شبکه با ورودی ۲ و ۴ نرون و یک لایه مخفی به‌عنوان شبکه برتر انتخاب شد که مقادیر ضریب همبستگی، RMSE و MAE برای این شبکه به‌ترتیب ۳۲/۱۵، ۱/۱۲ و ۰/۷۶ به‌دست آمد. در مدل SVM شبکه با ورودی ۳ به‌عنوان شبکه برتر انتخاب شد. مقدار پارامتر گاما برای این مدل ۰/۲۶۷۹۴۳۴ و مقدار کاست نیز ۱/۲۳۱۱۴۴ و مقادیر ضریب همبستگی، RMSE و MAE برای این شبکه به‌ترتیب ۵۳/۹۴، ۰/۹۳ و ۰/۶۵ به‌دست آمد.

در روش ترکیبی ۱، همان‌گونه که بیان شد، قسمت خطی به‌وسیله مدل ARIMA مدل‌سازی شده و سپس باقی‌مانده‌های مرحله قبل به‌وسیله مدل SVM مدل‌سازی شد. در روش ترکیبی ۲، خروجی نهایی حاصل، خروجی‌های وزن‌دار شده سه مدل ARIMA، شبکه عصبی مصنوعی و SVM می‌باشد. در نهایت، در روش ترکیبی ۳، مقادیر خروجی هر سه روش سری‌های زمانی، شبکه عصبی مصنوعی و

www.SID.ir

Archive of SID

SVM، به‌عنوان ورودی مدل SVM در نظر گرفته شد. در شکل‌های ۴ و ۵، هیدروگراف و دیاگرام پراکنش برای مدل برتر در داده‌های تست آورده شده است.

در این تحقیق، مقادیر دبی ماهانه ورودی به سد بوستان با استفاده از داده‌های ایستگاه هیدرومتری تمر و با استفاده از روش‌های سری‌های زمانی، شبکه عصبی، SVM و روش‌های ترکیبی پیش‌بینی شد. در نهایت برای مدل‌های سری زمانی مدل ARIMA (1,0,1) (2,0,0)، به‌عنوان مدل مناسب انتخاب شد.

برای روش شبکه عصبی مصنوعی پنج ورودی با تعداد ۱۰ نرون مورد بررسی قرار گرفت و در نهایت در این قسمت نیز با استفاده از معیارهای برازش بهترین ورودی‌ها و ساختار مدل های شبکه عصبی مصنوعی انتخاب شد. با توجه به نتایج به‌دست آمده، شبکه با ورودی ۲ و ۴ نرون به‌عنوان شبکه برتر انتخاب شد که مقادیر ضریب همبستگی، RMSE و MAE برای این شبکه به‌ترتیب ۳۲/۱۵، ۱/۱۲ و ۰/۷۶ متر مکعب بر ثانیه به‌دست آمد.

جدول ۲- مدل‌های مختلف سری‌های زمانی برای ایستگاه تمر

مدل	φ_1	φ_2	φ	θ_1	θ_2	θ_3	Φ_1	Φ_2	Φ_3	Θ_1	معیار آکاییک	معیار شوارتز
ARIMA(۲,۰,۱)(۲,۰,۱)	-۰/۰۸۵۱	۰/۲۸۲۹		-۰/۴۱۷۶			۱/۰۸۹۱	-۰/۲۱۲۶	۰/۱۱۸۶	۰/۹۱۶۴	-۱۳۳۶/۸۱	-۱۳۰۷/۹۵
ARIMA(۱,۰,۱)(۲,۰,۱)	۰/۵۸۶۶			۰/۲۵۲۸			۱/۱۱۰۶	-۰/۱۱۳۴		۰/۹۴۳۷	-۱۳۳۳۵/۵۵	-۱۳۱۴/۹۳
ARIMA(۱,۰,۱)(۱,۰,۱)	۰/۵۷۸۸			۰/۲۴۲۹			۰/۹۹۷۹			۰/۹۴۱۶	-۱۳۳۴/۰۷	-۱۳۱۷/۵۸
ARIMA(۱,۰,۲)(۱,۰,۱)	۰/۲۴۶۳			-۰/۰۸۸۶	-۰/۱۶۰۴		۰/۹۹۶۷			۰/۹۳۵۲	-۱۳۳۳/۸۱	-۱۳۱۳/۱۹
ARIMA(۱,۰,۰)(۱,۰,۱)	۰/۳۸۹۱						۰/۹۷۷۵			۰/۸۰۸۸	-۱۳۰۰/۳۱	۱۲۸۷/۹۴
ARIMA(۲,۰,۰)(۱,۰,۱)	۰/۳۲۳۹	۰/۱۱۹۶					۰/۹۹۶۷			۰/۹۳۵	۱۳۳۵/۸۱	-۱۳۱۹/۳۲
ARIMA(۲,۰,۱)(۲,۰,۱)	-۰/۰۵۲۶	۰/۲۶۷۲		-۰/۳۸۴۱			۱/۰۸۹۵	-۰/۰۹۲۶		۰/۹۲۴۷	-۱۳۳۲/۶۸	-۱۳۰۷/۹۴
ARIMA(۱,۰,۰)(۲,۰,۱)	۰/۳۶۶۸						۱/۰۸۳۲	-۰/۰۸۶		۰/۹۴۳۲	-۱۳۳۳/۲۱	-۱۳۱۶/۷۲
ARIMA(۱,۰,۰)(۱,۱,۱)	۰/۳۶۷۴						۰/۱۰۲۶			۰/۹۴۷۸	-۱۳۲۷/۵	-۱۳۱۵/۱۳
ARIMA(۲,۰,۰)(۱,۱,۱)	۰/۳۲۵	۰/۱۲۲۶					۰/۰۹۷۷			۰/۹۳۰۵	-۱۳۲۸/۰۵	۱۳۱۱/۵۶
ARIMA(۲,۰,۱)(۱,۰,۱)	۰/۳۲۳۹	۰/۱۱۹۶					۰/۹۹۶۷			۰/۹۳۵	-۱۳۳۵/۸۱	-۱۳۱۹/۳۲
ARIMA(۳,۰,۰)(۱,۱,۱)	۰/۳۳۰۶	۰/۱۴۱۲	-۰/۰۵۶۷				۰/۰۹۰۹			۰/۹۲۹۶	-۱۳۲۶/۰۵	۱۳۰۵/۴۴
ARIMA(۱,۰,۱)(۱,۱,۱)	۰/۵۷۲۹			۰/۲۳۳۵			۰/۱۰۰۹			۰/۹۴۷۲	-۱۳۲۹/۷۷	-۱۳۱۳/۲۸
ARIMA(۱,۰,۲)(۱,۱,۱)	۰/۲۹۲۳			-۰/۰۴۱۶	-۰/۱۴۷۳		۰/۰۹۱۲			۰/۹۳۰۲	-۱۳۲۶/۰۵	۱۳۰۵/۴۴
ARIMA(۱,۰,۳)(۱,۱,۱)	-۰/۹۲۶			-۱/۲۵۰۹	-۰/۵۲۶۴	-۰/۲۰۲۴	۰/۰۷۸۵			۰/۹۲۹۵	-۱۳۲۱/۵	-۱۲۹۶/۷۷
ARIMA(۲,۰,۱)(۱,۱,۱)	-۰/۰۱۷۱	۰/۲۵۳۷		-۰/۳۴۶۸			۰/۰۸۹۹			۰/۹۲۸۳	-۱۳۲۶/۹۱	۱۳۰۶/۳
ARIMA(۳,۰,۱)(۱,۱,۱)	-۰/۰۳۲۶	۰/۲۵۸۷	-۰/۰۱۳۴	-۰/۳۶۴۳			۰/۰۸۹۳			۰/۹۲۹۶	۱۳۲۴/۰۵	۱۲۹۹/۳۲
ARIMA(۱,۱,۱)(۱,۰,۱)	۰/۳۴۹۳			۰/۹۶۶۶			۰/۹۹۴۵			۰/۹۰۸۲	-۱۳۱۷/۹۲	۱۳۰۱/۴۳
ARIMA(۱,۱,۱)(۲,۰,۱)	۰/۳۴۱۵			۰/۹۵۷۴			۱/۰۴۵۵	-۰/۰۵۳۳		۰/۸۹۲۴	-۱۳۱۴/۲۵	-۱۲۹۳/۶۴
ARIMA(۱,۱,۱)(۲,۱,۱)	۰/۳۴۸			۰/۹۵۶			۰/۰۳۹۷	-۰/۱۴۱		۰/۸۷۱۶	-۱۳۰۷/۶۴	-۱۲۰۸۷/۰۳

φ_1: ضریب خود همبستگی غیر فصلی مرتبه۱ ، همبستگی غیر فصلی مرتبه۲

θ_1: ضریب میانگین متحرک غیرفصلی مرتبه۱

θ_2: ضریب میانگین متحرک غیرفصلی مرتبه۲

Φ_1: ضریب خود همبستگی فصلی مرتبه۱

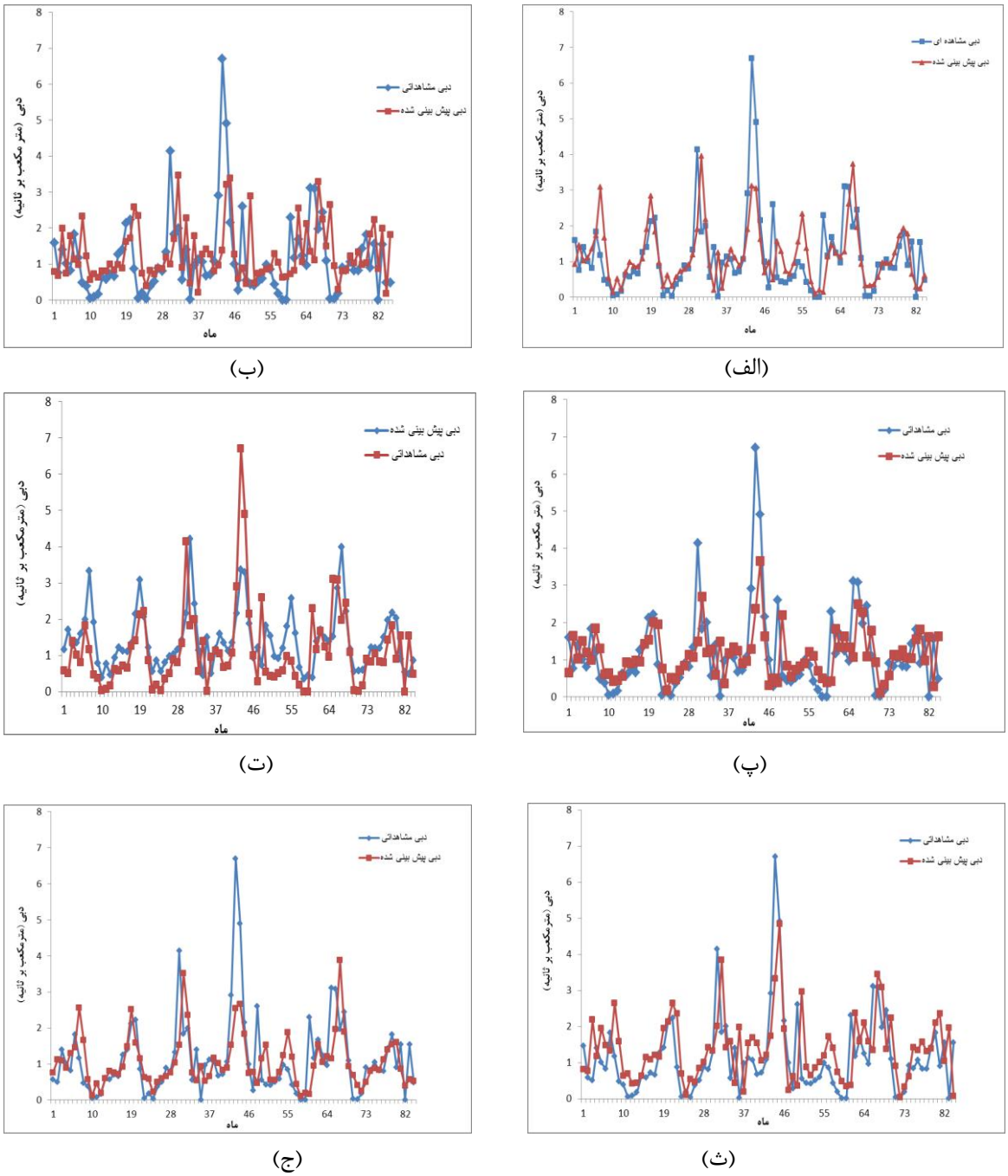
Φ: ضریب خود همبستگی فصلی مرتبه۲

Φ_3: ضریب خود همبستگی فصلی مرتبه۳

Θ_1: ضریب میانگین متحرک فصلی مرتبه۱

www.SID.ir

Archive of SID



شکل ۴- مقادیر دبی مشاهداتی و پیش‌بینی شده با استفاده از مدل‌های سری زمانی (الف)، شبکه عصبی مصنوعی (ب)، SVM (پ)، روش ترکیبی ۱ (ت)، روش ترکیبی ۲ (ث) و روش ترکیبی ۳ (ج)

برای مقایسه عملکرد مدل‌های مختلف از هفت معیار ارزیابی استفاده شد که نتایج آن در جدول ۳ آورده شده است.

در مدل SVM شبکه با ورودی ۳ به‌عنوان شبکه برتر انتخاب شد که مقادیر ضریب همبستگی، RMSE و MAE برای این شبکه به‌ترتیب ۵۳/۹۴، ۰/۹۳ و ۰/۶۵ به‌دست آمد.

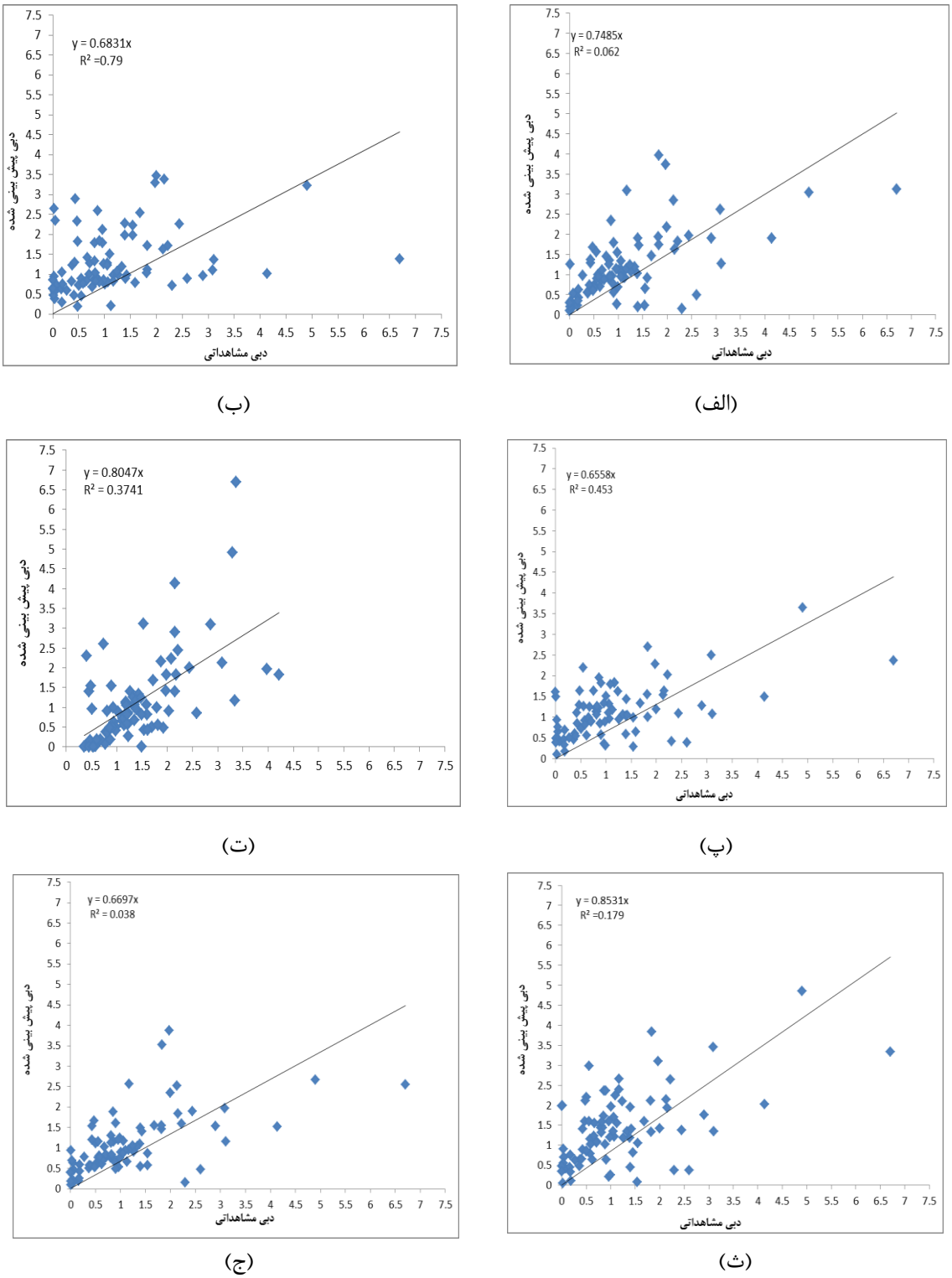
همان‌گونه که در شکل ۴ ملاحظه می‌شود، اکثر مدل‌ها مقادیر دبی کمینه را بهتر از مقادیر دبی بیشینه پیش‌بینی کرده‌اند.

با توجه به جدول ۳، که نتایج نهایی مدل‌های مختلف برای ایستگاه تمر را نشان می‌دهد، مدل ARIMA بهترین عملکرد را داشته است و با توجه به چهار معیار RMSE، AARE، MBE و CE بهترین مدل برای این ایستگاه می‌باشد. با توجه به معیار ضریب همبستگی مدل ARIMA بعد از مدل ترکیبی

www.SID.ir

Archive of SID

۱ بهترین مدل است و با توجه به معیار MAE مدل ترکیبی ۳، مدل برتر بوده و بعد از آن مدل ARIMA بهترین مدل بوده است.



شکل ۵- دیاگرام پراکنش مقادیر دبی مشاهده‌ای و پیش‌بینی شده با استفاده از مدل‌های سری زمانی (الف)، شبکه عصبی مصنوعی (ب)، SVM (پ)، روش ترکیبی ۱ (ت)، روش ترکیبی ۲ (ث) و روش ترکیبی ۳ (ج)

www.SID.ir

Archive of SID

جدول۳- نتایج معیارهای ارزیابی برای مدل‌های مختلف در ایستگاه تمر

مدل		ضریب همبستگی	RMSE (متر مکعب بر ثانیه)	MAE (متر مکعب بر ثانیه)	AARE	MBE	CE	PE
ARIMA	آموزش	۵۳/۸۳	۱/۴۹	۰/۷۱	۳/۰۱	۰/۲۵	۰/۲۷	۰/۶۴
	تست	۶۱/۰۵	۰/۸۸	۰/۵۹	۴/۷۱	-۰/۰۲۴	۰/۳۶	۰/۴۱
شبکه عصبی مصنوعی	آموزش	۲۱/۳۴	۱/۷۹	۱/۰۵	۱۲/۲۸	۰/۰۹	-۰/۰۶	۰/۵۲
	تست	۳۲/۱۵	۱/۱۲	۰/۷۶	۱۱/۶۱	-۰/۱۱۵	-۰/۰۴	۰/۴۸
SVM	آموزش	۵۴/۱۹	۱/۵۱	۰/۷۱	۴/۳۳	۰/۳۴	۰/۲۴	۰/۷۱
	تست	۵۳/۹۴	۰/۹۳	۰/۶۵	۱۴/۷۴	۰/۰۲۶	۰/۲۹	۰/۴۶
ترکیبی ۱	آموزش	۹۴/۷۱	۰/۶۲	۰/۱۹	۱/۷۱	۰/۰۰۴	۰/۸۸	۰/۴۳
	تست	۶۱/۲۰	۰/۹۳	۰/۷۰	۸/۳۸	-۰/۲۹۰	۰/۲۹	۰/۳۷
ترکیبی ۲	آموزش	۹/۲۷	۲/۰۶	۱/۴۳	۱۳/۲۰	۰/۰۰۱	-۰/۴۰	۰/۶۳
	تست	۵۵/۸۵	۰/۹۹	۰/۷۴	۱۷/۴۹	-۰/۲۴	۰/۲۰	۰/۲۸
ترکیبی ۳	آموزش	۶۶/۳۴	۱/۳۲	۰/۵۵	۲/۶۴	۰/۲۲	۰/۴۲	۰/۲۷
	تست	۵۹/۱۱	۰/۸۹	۰/۵۶	۵/۱۶	۰/۰۷	۰/۳۴	۰/۴۲

بر اساس معیار PE هم مدل ترکیبی ۲ بهترین عملکرد را داشته است. بنابراین، مدل ARIMA بهترین مدل برای ایستگاه تمر می‌باشد و مدل سری زمانی نسبت به مدل شبکه عصبی مصنوعی و SVM عملکرد بهتری داشته است که این موضوع با نتایج Kafil و Ahmadi (۲۰۰۶) مطابقت دارد. آن‌ها بیان کردند که روش شبکه عصبی مصنوعی برای پیش‌بینی‌های بلند مدت و روش سری زمانی برای پیش‌بینی‌های کوتاه مدت مناسب‌تر می‌باشند. ولی با نتایج Castellan-Mendez (۲۰۰۴)، Valipour و همکاران (۲۰۱۳) و Razavi و همکاران (۲۰۰۶) مطابقت ندارد. در این پژوهش، فقط از مقادیر گذشته دبی به‌عنوان ورودی مدل‌های مختلف استفاده شده است. در این حالت، در مدل‌های سری زمانی از ساختار همبستگی موجود در داده‌ها در مدل‌سازی استفاده می‌کنند. همچنین، در این مدل‌ها الگوریتم مدون و مشخصی برای تعیین ساختار بهینه مدل وجود دارد، مسئله‌ای که کمتر در مدل‌های غیر خطی هوش مصنوعی مشاهده می‌شود. چنانچه از مقادیر متغیرهایی دیگر به‌عنوان ورودی مدل استفاده می‌شد، در این حالت ممکن بود، روش‌های هوش مصنوعی عملکرد بهتری داشته باشند. Wang و همکاران (۲۰۰۵) نیز به نتایج مشابهی برای شبکه عصبی و مدل‌های سری‌های زمانی در پیش‌بینی دبی رودخانه زرد چین رسیدند. در پژوهش Nayak و همکاران (۲۰۰۴) در پیش‌بینی دبی رودخانه بایتارانی به نتایج متفاوت رسیدند که البته این مسئله را می‌توان در این دید که در آن تحقیق از مدل‌های ARMA برای پیش‌بینی جریان رودخانه استفاده شده و هیچ‌گونه اشاره‌ای به استانداردسازی فصلی و یا استفاده از مدل‌های ARIMA که در این پژوهش بهترین جواب را داده، نشده است. مشابه همین مسئله را می‌توان در تحقیقات، Jin و همکاران (۱۹۹۹) در پیش‌بینی دبی ورودی به سد حوضه اورسیای هند، Kisi و همکاران (۲۰۰۴) در پیش‌بینی دبی رودخانه گوکشودر و Jain و Kumar (۲۰۰۷) در پیش‌بینی مقادیر مقادیر دبی رودخانه کلرادو دید که در دو پژوهش آخر برای مقایسه روش‌های سری‌های زمانی و هوش مصنوعی فقط از مدل‌های AR استفاده شده است و حتی جزء میانگین متحرک نیز در این مدل‌های مشاهده نمی‌شود. در پژوهش Huang و همکاران (۲۰۰۴) اگر چه از مدل‌های ARIMA استفاده شده است، ولی برای شبکه عصبی از مقادیر بارش نیز به‌عنوان ورودی کمکی استفاده شده است. همچنین، مدل شبکه عصبی مصنوعی با توجه به اکثر معیارهای ارزیابی بدترین عملکرد را در بین شش مدل داشته است. در حالت ترکیبی که سه روش استفاده شد، ملاحظه می‌شود که این روش‌های ترکیبی باعث بهبود عملکرد

www.SID.ir

Archive of SID

مدل‌های شبکه عصبی مصنوعی و SVM به‌صورت منفرد می‌شود که با نتایج Zhang (۲۰۰۳)، Cheng و همکاران (۲۰۱۵) و Young و همکاران (۲۰۱۷) مطابقت دارد.

منابع مورد استفاده

1. Azizi, Gh. 2005. Investigation of drought and wet years and possible to predict them using time-series model in Hormozgan Station. Journal of Geographical Research, 79: 48-63.
2. Castellano-Mendez, M., W. Gonzalez-Manteiga, M. Febrero-Bande, J.M. Prada-Sanchez and R. Lozano-Calderon. 2004. Modelling of monthly and daily behavior of the run off the Xallas River using Box-Jenkins and Neural networks methods. Journal of Hydrology, 296: 38-58.
3. Chen, S.T. and P.S. Yu. 2007. Pruning of support vector networks on flood forecasting. Journal of Hydrology, 347: 67–78.
4. Cheng, C.T., Z.K. Feng, W.J. Niu and S.L. Liao. 2015. Heuristic methods for reservoir monthly inflow forecasting, a case study of Xinfengjiang Reservoir in Pearl River China. Journal of Water, 7: 4477-4495
5. Young, C.C., W.C. Liu and M.C. Wu. 2017. A physically based and machine learning hybrid approach for accurate rainfall-runoff modeling during extreme typhoon events. Applied Soft Computing, 53: 205-216.
6. Dibike, Y.B., S., Velickov, D., Sololatine and M.B. Abbott. 2001. Model induction with support vector machine introduction and application. Journal of Computing in Civil Engineering, 15: 208–216.
7. El-Shafie, A., M. Reda Taha and A. Noureldin. 2007. A neuro-fuzzy model for inflow forecasting of the Nile River at Aswan High Dam. Water Resources Management, 21: 533–556.
8. Estaki, A., A. Mosaedi, A. Zarei and M. Mesdaghi. 2016. Inflow forecasting and studying with time series. 2nd Iranian National Irrigation and Drainage Congress, Isfahan University of Technology, Isfahan (in Persian).
9. Firat, M. and M. Gungor. 2007. River flow estimation using adaptive neuro fuzzy inference system. Mathematics and Computers in Simulation, 75(3-4): 87-96.
10. Gorbani, Kh., E. Sohrabian and M. Salari Jazi. 2016. Evaluation of hydrological and data mining models in monthly river discharge simulation and prediction, case study: Araz-Kouseh Watershed. Journal of Water and Soil Conversation, 23(1): 203-217 (in Persian).
11. Hasanah, Y. and M. Herlina. 2013. Flood prediction using transfer function model of water discharge approach in Katulampa Dam. Procedia Environmental Sciences, 17(0): 317-326.
12. Hornik, K., M. Stinchcombe and H. White. 1989. Multilayer feedforward networks are universal approximates. Neural Networks, 2(5): 359-366.
13. Huang, w., B. Xu and A. Chan-Hilton. 2004. Forecasting flows in Apalachicola River using neural networks. Hydrological Processes. 18(13): 2545–2564.
14. Jain, A. and A.M. Kumar. 2007. Hybrid neural network models for hydrologic time series forecasting. Applied Soft Computing Journal, 7(2): 585- 592.
15. Jain, S.K., A. Dasand and D.K. Srivastava. 1999. Application of ANN for reservoir inflow prediction and operation. Journal of Water Resources Planning and Management, 125(5): 263-271.
16. Jalal Kamali, N. 2002. Monthly flow prediction into the Jiroft Reservoir Dam by the theory of time series. The Sixth Congress of International River Engineering, Shahid Chamran University, Ahvaz (in Persian).
17. Kafil, M. and M. Ziatabar Ahmadi. 2006. Comparison of artificial neural network and time series methods in forecasting river flow rate. The Second Conference of Iran Water Resources Management, Isfahan University of technology, 189 pages (in Persian).
18. Karamouz, M. and Sh. Araghinejd. 2005. Advanced hydrology. Amirkabir University of Technology Press, 256 pages (in Persian).
19. Kisi, O. 2004. River flow modeling using artificial neural networks. Journal of Hydrologic Engineering, 9(1): 60-63.
20. Mianabadi, H. and A. Afshar. 1999. Application of periodic monthly time series in monthly runoff production and integrated management of water resources in Kashaf-roud Basin. Third Conference on Water Resources Management, Tabriz, Iran (in Persian).
21. Nayak, P.C., K.P. Sudheer, D.M. Rangan and K.S. Ramasastri. 2004. A neuro-fuzzy computing technique for modeling hydrological time series. Journal of Hydrology, 291: 52–66.
22. Niroumand, H. and A. Bozorgnia. 1993. Introduction to time series. Ferdowsi University of Mashhad

www.SID.ir

Archive of SID

Press, 289 pages (in Persian).

23. Razavi, S., Sh. Araghinejad, H. Maneshi and N. Hosseini. 2006. Modeling of inflow to the reservoir using recurrent neural network with the approach of forget operating. The Second Conference of Iran Water Resources Management, Isfahan University of technology, 189 pages (in Persian).
24. Rezaei, A. 2004. Regional modeling of peak discharges in sub-watershed of Sefidroud Dam using artificial neural network. PhD Thesis, Tehran University, 146 pages (in Persian).
25. Sadoddin, A., A. Mosaedi and M. Halili. 2009. Management plan for reservoir systems: feasibility study on water resources management using Decision Support System (DSS), case study: Gorgan-rud River. Applied Research Plan, Golestan Regional Water Company (in Persian).
26. Salamatian, S., A. Dehghani and M. Aghdasian. 2006. Smart estimating of river discharge using neural network. The First Regional Conference on Optimum Utilization of Karoon and Zayandehrud Basins Water Resources. University of Shahrekord (in Persian).
27. Valipour, M., M.E. Banihabib and S.M.R. Behbahani. 2013. Comparison of the ARMA, ARIMA, and the autoregressive artificial neural network model in forecasting the monthly inflow of Dez Dam Reservoir. Journal of Hydrology, 476: 433–441.
28. Van Dijk, A.I.J.M., L.A. Bruijnzeel and E.H. Eisma. 2003. A methodology to study rain splash and wash processes under natural rainfall. Hydrological Processes, 17: 153–167.
29. Wang, W., P.V. Gelder and J.K. Vrijling. 2005. Improving daily stream flow forecasting by combining ARMA and ANN models. Journal of Hydro-Environment Research, 18: 12-24.
30. Zhang, G.P. 2003. Time series forecasting using a hybrid ARIMA. Neurocomputing, 50: 159–175.

www.SID.ir