

مدل‌سازی دبی ماهانه ورودی به مخزن سد جامیشان با مدل‌های خودهمبسته با میانگین متحرک تجمعی و سامانه استنتاج فازی - عصبی انطباقی

حمید معینی^{1*}، حسین بنکداری²، سید احسان فاطمی³، عیسی ابتهاج⁴

تاریخ دریافت: 94/04/03 تاریخ پذیرش: 94/10/29

¹- دانشجوی کارشناسی ارشد مهندسی عمران-آب دانشگاه رازی کرمانشاه

²- استاد گروه عمران دانشگاه رازی، کرمانشاه

³- استادیار گروه منابع آب، دانشگاه رازی، کرمانشاه

⁴- دانشجوی دکتری مهندسی عمران-آب و سازه‌های هیدرولیکی، دانشگاه رازی کرمانشاه

*مسئول مکاتبات، پست الکترونیکی: h.moeeni68@gmail.com

چکیده

مدل‌سازی سری‌های زمانی هیدرولوژیکی یکی از مهم‌ترین موضوعات در مدیریت منابع آب است. در این پژوهش دبی ماهانه ورودی به مخزن سد جامیشان در استان کرمانشاه (غرب ایران) با استفاده از مدل‌های خودهمبسته با میانگین متحرک تجمعی (ARIMA) و سامانه استنتاج فازی-عصبی انطباقی (ANFIS) مدل‌سازی گردید. این مدل‌ها به ترتیب بر پایه روش‌های تصادفی (Stochastic) و هوش مصنوعی (AI) هستند. برای مدل‌سازی تا بیشینه پنج پارامتر برای مدل ARIMA در نظر گرفته شد که 1296 مدل تولید و بر سری زمانی برازش داده شد. برای مدل ANFIS نیز 14 نوع ترکیب ورودی با استفاده از دبی‌های با تأخیرهای مختلف تعیین گردید. در تولید سامانه استنتاج فازی (FIS) از دو روش شبکه‌بندی (GP) و خوشه‌بندی تفریقی (SC) استفاده شد. در آموزش شبکه نیز دو الگوریتم پس‌انتشار خطا (BP) و الگوریتم ترکیبی (Hybrid) به‌کاربرده شد. دبی‌های ماهانه مدل‌سازی‌شده توسط مدل‌های ARIMA و ANFIS به‌کمک شاخص‌های مختلف از جمله شاخص میانگین قدر مطلق خطای نسبی (MARE) با یکدیگر مقایسه شدند که مقدار آن به ترتیب برابر 0/398 و 0/8 برای هر مدل به‌دست آمد. نتایج به‌دست آمده نشان داد مدل ARIMA در مدل‌سازی دبی‌های با مقادیر کم و مدل‌سازی‌های کوتاه‌مدت و درازمدت بسیار توانمندتر از ANFIS هست.

واژه‌های کلیدی: استوکستیک، دبی، مدل‌سازی، ANFIS، ARIMA

Modeling the Monthly Inflow to Jamishan Dam Reservoir Using Autoregressive Integrated Moving Average and Adaptive Neuro- Fuzzy Inference System Models

H Moeeni^{1*}, H Bonakdari², SE Fatemi³, I Ebtehaj⁴

Received: 24 June 2015 Accepted: 19 January 2016

¹- M.Sc. Student of Civil-Water Engineering, Kermanshah Razi University, Iran

²- Assoc. Prof., Dept. of Civil Engineering, Razi University, Iran

³- Assist. Prof., Dept. of Water Resource, Kermanshah Razi University, Iran

⁴- Ph.D. Candidate, Civil-Water and Hydraulic Structures, Kermanshah Razi University, Iran

* Corresponding Author, Email: h.moeeni68@gmail.com

Abstract

Hydrological time series modeling is one of the most important issues in water resource management. In this paper monthly inflow to Jamishan dam reservoir in Kermanshah province (west of Iran) is modeled by AutoRegressive Integrated Moving Average (ARIMA) and Adaptive Neuro-Fuzzy Inference System (ANFIS) models. These models are based on stochastic and Artificial Intelligence (AI) methods, respectively. For modeling up to five parameters in the ARIMA model were used and produced 1296 models which were fitted on the time series. In ANFIS model 14 input combinations were defined using the discharges with different lags. Two states of Grid Partitioning (GP) and Subtractive Clustering (SC) were used in Fuzzy Interface System (FIS) generation. Also, in training network Back Propagation (BP) and hybrid algorithms were used. Monthly modeled discharges were compared in the ARIMA and ANFIS models by some indexes such as Mean Absolute Relative Error (MARE) index which was obtained 0.398 and 0.8 for each model, respectively. The result showed that the ARIMA model was much more accurate than ANFIS model in modeling low discharges and also in short and long times modeling.

Keywords: ANFIS, ARIMA, Inflow, Modeling, Stochastic

می‌توان به مدل‌های خودهمبسته با میانگین متحرک
جمعی¹ (ARIMA) و سامانه استنتاج فازی-عصبی
انطباقی² (ANFIS) اشاره کرد. مدل ARIMA یکی از

مقدمه

سری‌های زمانی هیدرولوژیکی ابزاری بسیار
کارآمد برای شناخت ماهیت پدیده‌های هیدرولوژیکی
هستند که با داشتن شناخت کافی از آنها می‌توان
تغییرات آینده را مدل‌سازی و پیش‌بینی کرد. روش‌های
مدل‌سازی سری‌های زمانی، متنوع هستند. از این بین

¹ Auto regressive integrated moving average (ARIMA)

² Adaptive network based fuzzy inference system (ANFIS)

ANFIS بهتر بود. کیسی و همکاران (2012) به مدل‌سازی بار رسوب رودخانه‌ی کامبرلند¹¹ در کشور آمریکا با استفاده از داده‌های 1979 تا 1989 پرداختند. بدین منظور روش‌های ANN، ANFIS، بردار ماشین پشتیبان¹² (SVM) و GEP را به‌کار بردند. ایشان از داده‌های دبی و رسوب به‌عنوان ورودی‌های مدل‌ها استفاده کردند. نتیجه بیان‌گر قابلیت بیشتر مدل GEP بود.

از آنجایی‌که هریک از روش‌های تصادفی و هوش محاسباتی مزایا و معایبی دارند مقایسه آن‌ها با یکدیگر نیز از جمله موضوعات مورد مطالعه بوده است (احمدی و همکاران 1394 ب، خطیبی و همکاران 2014). خلقی و همکاران (1385) سری زمانی دبی کمینه ماهانه (کم‌آبی-های ماهانه) یک، سه و هفت روزه رودخانه را با مدل‌های ARIMA و ANFIS پیش‌بینی کردند. ایشان بیشینه سه پارامتر برای مدل ARIMA در نظر گرفتند. نتایج نشان داد مدل ARIMA از توانایی و دقت بیشتری برخوردار است. فیرات (2008) به مقایسه کارایی و ظرفیت مدل‌های ANFIS، ANN و AR در مدل‌سازی دبی روزانه در دو رودخانه سیحان¹³ و ساین¹⁴ در ترکیه پرداخت. ایشان بدین‌منظور به‌ترتیب از دوره‌های آماری 2000-1986 و 2000-1992 برای رودخانه‌ها استفاده کرد. نتایج بیان‌گر برتری ANFIS نسبت به ANN و AR بود. نواز و همکاران (2015) با استفاده از مدل ANFIS و مدل خودهمبسته با ورودی بیرونی¹⁵ (ARX) به مدل‌سازی سطح رودخانه بیکوک¹⁶ در حوضه‌ای گرمسیری در مالزی پرداخت. نتایج بیان‌گر برتری مدل ANFIS در برآورد سطح رودخانه در نقاط اوج بود.

مدل‌های تصادفی³ است که مبتنی بر نظریه آمار و احتمالات هست. ANFIS نیز یکی از مدل‌های نروفازی⁴ است که بر پایه هوش مصنوعی⁵ هست.

استفاده از مدل‌های تصادفی در هیدرولوژی بسیار گسترش یافته است به‌طوری‌که محققان زیادی برای مدل‌سازی و پیش‌بینی سری‌های زمانی هیدرولوژیکی از این روش‌ها استفاده کردند (ناوه و همکاران 1391، ولی‌پور 2012). احمدی و همکاران (1393) به مدل‌سازی دبی ماهانه و روزانه رودخانه باراندوزچای ارومیه در دوره آماری 1352 تا 1388 با استفاده از مدل خطی خودهمبسته با میانگین متحرک⁶ (ARMA) و مدل غیرخطی دوخطی⁷ (BL) پرداختند. نتایج بیان‌گر بهبود دقت مدل‌سازی در مدل BL بود. ولی‌پور (2015) توانایی مدل‌های ARIMA و خودهمبسته با میانگین متحرک تجمعی فصلی⁸ (SARIMA) را برای پیش‌بینی رواناب درازمدت برای داده‌های تمام ایالات آمریکا در دوره‌های آماری مختلف مطالعه کرد. نتایج نشان داد که برای تمام ایستگاه‌های مورد بررسی مدل SARIMA بهتر است به‌طوری‌که برای این مدل خطای نسبی برای تمام ایالت‌ها کمتر از 5 درصد بود.

هوش محاسباتی⁹ (CI) نیز در مهندسی آب و هیدرولوژی کاربرد زیادی پیدا کرده و مورد استفاده قرار گرفته است (کیسی و سیگیزقلو 2007، ابتهاج و بنکداری 2014). احمدی و همکاران (1394 الف) به برآورد تبخیر و تعرق ماهانه با استفاده از داده‌های 38 ساله‌ی شش ایستگاه در شمال‌غرب ایران به‌کمک روش‌های ANFIS و برنامه‌ریزی ژنتیک¹⁰ (GEP) پرداختند. هرچند هر دو مدل عملکرد قابل‌قبولی داشتند اما مدل

³ Stochastic method

⁴ Neuro-fuzzy

⁵ Artificial intelligence (AI)

⁶ Auto regressive moving average (ARMA)

⁷ Bilinear model (BL)

⁸ Seasonal auto regressive integrated moving average

⁹ Computational intelligence (CI)

¹⁰ Genetic programming (GEP)

¹¹ Cumberland

¹² Support vector machine (SVM)

¹³ Seyhan river

¹⁴ Cine river

¹⁵ Autoregressive with exogenous input (ARX)

¹⁶ Bekok river

دربریگیرنده قابلیت‌های هر دو روش شبکه‌های عصبی³ و و سامانه استنتاج فازی است.

برای مدل‌سازی سری‌های زمانی هیدرولوژیکی طول سری باید به‌گونه‌ای باشد که کلیه دوره‌های خشک‌سالی و ترسالی را شامل باشد. با استفاده از ضریب هرست⁴ (هرست و همکاران 1965) طول داده‌ها موردبررسی قرار گرفت. اگر این ضریب بیشتر از 0/5 باشد یعنی سری دارای طول کافی است که در این مطالعه برابر 0/711 به‌دست آمد. در مدل ARIMA داده‌ها باید از توزیع نرمال پیروی کنند (سالاس و همکاران 1980). به‌منظور نرمال کردن داده‌ها و ایستاسازی انحراف معیار از تبدیل باکس-کاکس⁵ (باکس و کاکس 1964) استفاده شد. به‌کمک آزمون جارکو-برا⁶ (جارکو و برا 1980) نرمال شدن داده‌های دبی به‌صورت زیر بررسی گردید:

$$J_b = n \left[\frac{\sum_{k=1}^n s_k^2}{6} + \frac{(K_u - 3)^2}{24} \right] \frac{\bar{\sigma}}{\sigma} \quad [1]$$

که در آن J_b آماره آزمون، s_k چولگی، K_u کشیدگی و n تعداد داده‌ها است. در صورتی که مقدار احتمال متناظر با این آماره در توزیع مربع‌کای بیشتر از سطح اعتماد موردنظر باشد، سری زمانی نرمال خواهد بود.

در روش‌های تصادفی، داده‌ها باید به جزء روند⁷، روند⁷، جزء تناوب⁸، جزء خودهمبسته و جزء تصادفی تقسیم شوند. اجزاء روند و تناوب باعث نا ایستا شدن سری می‌شوند که باید از سری حذف شوند. در مدل ARIMA، اجزاء روند و تناوب به‌ترتیب با تفاضل‌گیری-های غیرفصلی و فصلی حذف می‌شوند. در نتیجه سری ایستا می‌شود. شکل 1 نمودار تابع خودهمبستگی⁹ (ACF)

در این پژوهش عملکرد مدل‌های ARIMA به‌عنوان یک روش مبتنی بر آمار و احتمالات و ANFIS به‌عنوان یک روش مبتنی بر هوش مصنوعی در مدل-سازی دبی ماهانه با یکدیگر مقایسه می‌شود. عملکرد این دو مدل در مدل‌سازی دبی‌های پایه (دبی‌های با مقادیر کم) و دبی‌های سیلابی (دبی‌های با مقادیر زیاد) به‌طور جداگانه سنجیده خواهد شد. در آخر نیز دقت استفاده از نتایج مدل‌ها در کوتاه‌مدت و بلندمدت بررسی خواهد شد. به‌این‌ترتیب تأثیر تغییر طول دوره اعتبارسنجی بر عملکرد مدل‌ها شناسایی می‌شود.

مواد و روش‌ها

داده‌ها و منطقه مورد مطالعه

در این پژوهش دبی ماهانه 30 سال آبی از اکتبر 1980 تا سپتامبر 2011 مورد مطالعه قرار گرفت. این داده‌ها از ایستگاه هیدرومتری پیرسلیمان با طول جغرافیایی $34^{\circ}14'$ و عرض جغرافیایی $46^{\circ}40'$ بر روی رودخانه جامیشان به‌دست‌آمده است. حوضه رودخانه جامیشان یکی از زیرحوضه‌های رودخانه گاماسیاب است که در نزدیکی شهرستان سنقر در استان کرمانشاه واقع است و سد جامیشان بر روی آن قرار دارد. این سد با حجم نرمال 62/8 میلیون مترمکعب با هدف تأمین آب زراعی منطقه و کنترل سیلاب و جلوگیری از خسارات ناشی از آن احداث می‌گردد.

پیش‌پردازش داده‌ها

در این پژوهش از مدل ARIMA و ANFIS به‌منظور مدل‌سازی دبی ماهانه استفاده شد. زیرا مدل ARIMA دربرگیرنده هر دو جزء خودهمبسته¹ و جزء تصادفی² داده‌ها در مدل‌سازی است. مدل ANFIS نیز

³ Neural networks (NNs)

⁴ Hurst coefficient

⁵ Box-Cox transformation

⁶ Jarque-Bera

⁷ Trend component

⁸ Periodical component

⁹ Autocorrelation function (ACF)

¹ Autoregressive component

² Random component

مدل خودهمبسته با میانگین متحرک تجمعی

مدل ARIMA به مدل‌سازی اجزاء خودهمبسته و تصادفی می‌پردازد و با $ARIMA(p,d,q)(P,D,Q)_\omega$ نشان داده می‌شود. این مدل دارای رابطه زیر است (سالاس و همکاران 1980):

$$\begin{aligned} & (1 - \Phi_1 B^\omega - \Phi_2 B^{2\omega} - \dots - \Phi_p B^{p\omega}) \\ & (1 - j_1 B - j_2 B^2 - \dots - j_p B^p)(1 - B)^\omega \\ & (1 - B)^d x_t = (1 - \Theta_1 B^\omega - \Theta_2 B^{2\omega} - \dots - \Theta_q B^{q\omega}) \varepsilon_t \end{aligned} \quad [2]$$

که در آن x_t دبی مشاهده‌شده، ε_t جزء تصادفی، ω دوره تناوب، B عملگر تفاضلی به صورت $B(x_t) = x_{t-1}$ ، D برابر $(1-B)^\omega$ برای D امین تفاضل فصلی با گام ω ، d برابر $(1-B)^d$ برای d امین تفاضل فصلی غیرفصلی، φ ، q ، P و Q مرتبه‌های خودهمبسته و میانگین متحرک فصلی و غیرفصلی و θ ، Θ پارامترهای خودهمبسته و میانگین متحرک فصلی و غیرفصلی هستند.

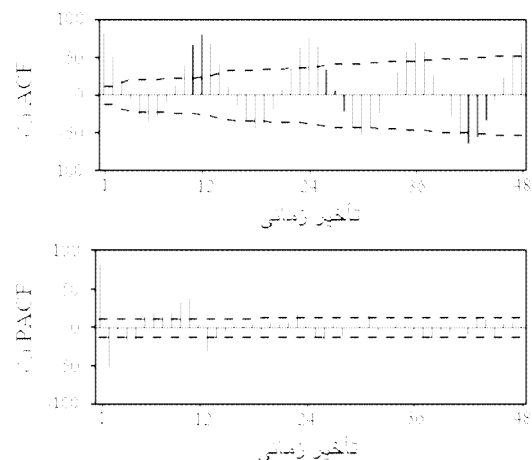
با در نظر گرفتن بیشینه پنج پارامتر از هر نوع خودهمبسته و میانگین متحرک، 1296 مدل ARIMA بر داده‌ها برازش داده شد. پس از برازش مدل‌ها، آن‌هایی قابل قبول خواهند بود که دارای پارامترهایی باشند که به‌طور معنی‌داری با صفر اختلاف دارند و همچنین دارای باقیمانده‌های مستقل باشند (کرایر و چان 2008). در این تحقیق از آزمون تی - استیودنت² برای بررسی معنی‌دار بودن پارامترها به‌صورت زیر استفاده شد:

$$\tau = \frac{Pr}{Se} \quad [3]$$

که در آن τ آماره آزمون، Pr مقدار برآورد شده‌ی هر پارامتر (اعم از θ ، Θ ، φ و Φ) و Se خطای معیار³ برآورد شده است. اگر سطح احتمال متناظر با آماره τ در توزیع t از 5 درصد کمتر باشد ($P_\tau < 0/05$)، یعنی پارامتر موردنظر معنی‌دار است.

(ACF) و خودهمبستگی جزئی¹ (PACF) مربوط به داده‌ها را نشان می‌دهد. نمودار ACF به صفر میل نمی‌کند و دارای دوره تناوب 12 هست. نمودار PACF بلافاصله پس از اولین گام تغییر علامت داده است و قطع می‌شود. پس برای ایستاسازی سری فقط به یکبار تفاضل‌گیری فصلی نیاز است. برای تعیین تعداد پارامترهای مدل ARIMA نیز از این نمودارها استفاده می‌شود. با توجه به ACF خودهمبستگی در تأخیرهای 1، 2، 3، 6، 12 و 24 دارای مقادیر زیادی است که تأثیر زیاد دبی‌های با تأخیرهای گفته‌شده بر یکدیگر را نشان می‌دهد. در نتیجه 14 نوع ورودی با ترکیب‌های مختلف برای مدل ANFIS در نظر گرفته شد. خروجی هر مدل دبی ماه جاری (Q_t) است که تابعی از دبی در ماه‌های قبل هست. این مدل‌ها به‌شرح زیر می‌باشند:

- ANFIS-Model 1: Q_{t-1}
- ANFIS-Model 2: Q_{t-1}, Q_{t-2}
- ANFIS-Model 3: Q_{t-1}, Q_{t-12}
- ANFIS-Model 4: Q_{t-1}, Q_{t-6}
- ANFIS-Model 5: $Q_{t-1}, Q_{t-6}, Q_{t-12}$
- ANFIS-Model 6: $Q_{t-1}, Q_{t-2}, Q_{t-12}$
- ANFIS-Model 7: $Q_{t-1}, Q_{t-2}, Q_{t-6}$
- ANFIS-Model 8: $Q_{t-1}, Q_{t-2}, Q_{t-3}$
- ANFIS-Model 9: $Q_{t-1}, Q_{t-2}, Q_{t-6}, Q_{t-12}$
- ANFIS-Model 10: $Q_{t-1}, Q_{t-2}, Q_{t-3}, Q_{t-12}$
- ANFIS-Model 11: $Q_{t-1}, Q_{t-2}, Q_{t-3}, Q_{t-6}$
- ANFIS-Model 12: $Q_{t-1}, Q_{t-2}, Q_{t-3}, Q_{t-6}, Q_{t-12}$
- ANFIS-Model 13: $Q_{t-1}, Q_{t-2}, Q_{t-3}, Q_{t-12}, Q_{t-24}$
- ANFIS-Model 14: $Q_{t-1}, Q_{t-2}, Q_{t-3}, Q_{t-6}, Q_{t-12}, Q_{t-24}$



شکل 1- نمودارهای ACF و PACF مربوط به دبی ماهانه.

¹ Partial autocorrelation function (PACF)

² t-student

³ Standard error (SE)

از آزمون باکس- پییرس¹ (باکس و پییرس 1970 و مکلوود 1978) برای بررسی مستقل بودن باقی-مانده‌های مدل‌های ARIMA به صورت زیر استفاده گردید:

$$Q^* = n(n+2) \sum_{l=1}^L \frac{r_l^2(\epsilon)}{n-1} \quad [4]$$

که در آن $r_l(\epsilon)$ ضریب خودهمبستگی باقیمانده‌ها با تأخیر l $n=N-d-D\omega$ تعداد ماه‌های باقی‌مانده پس از تفاضل‌گیری و L بیشینه زمان تأخیر در نظر گرفته شده است. $k=p+q+P+Q$ به عنوان تعداد پارامترهای مدل تعریف می‌شود. در صورتی که سطح احتمال متناظر با Q^* در توزیع مربع کای با درجه‌ی آزادی $DF=L-K-1$ از سطح معنی‌داری موردنظر یعنی 5 درصد بزرگتر باشد $(P_{Q^*} > 0/05)$ ، سری باقی‌مانده‌ها مستقل خواهد بود.

دوره‌نگار تجمعی² (هیپیل و مکلوود 1994) نیز برای بررسی حذف کامل جزء دوره در باقی‌مانده‌ها به کار برده شد. این آزمون که به نوعی مستقل بودن باقی‌مانده‌ها را بررسی می‌کند به صورت زیر است:

$$MSD(h_j) = \sum_{t=1}^n \xi_t^2 \cos(2\pi h_j t) + \sum_{t=1}^n \xi_t^2 \sin(2\pi h_j t) \quad [5]$$

$$h_j = j/n$$

$$P_i = \frac{2 \sum_{j=1}^i MSD(h_j)}{n^2 \sigma_\epsilon^2}, \quad i = 1, 2, \dots, \frac{n}{2} \quad [6]$$

که در آن P_i دوره‌نگار تجمعی باقی‌مانده‌ها، h_j فرکانس، MSD میانگین مربع انحرافات، ξ_t مقادیر باقی‌مانده‌ها در زمان t ، σ_ϵ^2 واریانس باقی‌مانده‌ها و n تعداد ماه‌های باقی‌مانده پس از تفاضل‌گیری است. اگر سری باقی‌مانده‌ها مستقل باشد، نمودار P_i نسبت به h_j نزدیک به خط واصل نقاط $(0,0)$ و $(0/5, 0)$ هست.

سامانه استنتاج فازی - عصبی انطباقی

مدل ANFIS یک روش ترکیبی از محاسبات نرم³ و شبکه‌های عصبی است. با در نظر گرفتن یک نمونه از این فرآیند برای یک سامانه استنتاج فازی با دو ورودی و یک خروجی، ساختار کلی و پنج لایه ANFIS در ادامه شرح داده می‌شود:

لایه اول (گره‌های ورودی): هر گره در این لایه درجات عضویت از یک متغیر ورودی را تولید می‌کند.

$$O_i^1 = \mu_{A_i}(Q_{t-1}) \quad i=1,2$$

$$O_i^1 = \mu_{B_{i-2}}(Q_{t-2}) \quad i=3,4 \quad [7]$$

که Q_{t-1} و Q_{t-2} ورودی‌های اول و دوم و μ_{A_i} و $\mu_{B_{i-2}}$ توابع عضویت مجموعه‌های فازی A_i و B_i هست. به منظور تعیین درجه‌های عضویت، توابع عضویت متفاوتی مثل مثلثی، ذوزنقه‌ای، گوسی و زنگوله‌ای وجود دارد. تابع گوسی دارای مزایای هموار و غیر صفر بودن است و همچنین پارامترهای کمتری دارد. به همین دلیل در این تحقیق از تابع عضویت گوسی استفاده گردید.

لایه دوم (گره‌های قوانین): در این لایه از عملگرهای فازی به منظور استخراج قوه انگیزش هر قانون فازی به مجموع قوه‌های انگیزش درجه‌های عضویت محاسبه شده در لایه اول به دست می‌آید:

$$O_i^2 = w_i = \mu_{A_i}(Q_{t-1}) \mu_{B_i}(Q_{t-2}) \quad i=1,2 \quad [8]$$

لایه سوم (گره‌های میانگین): در این لایه نسبت قوه انگیزش هر قانون فازی به مجموع قوه‌های انگیزش تمامی قوانین به صورت زیر تعیین می‌گردد:

$$O_i^3 = \bar{w}_i = \frac{w_i}{w_1 + w_2} \quad i=1,2 \quad [9]$$

³ Soft computing

¹ Box-Pierce

² Cumulative periodogram

که در آن Q_{ni} مقدار نرمال‌شده دبی مشاهده‌شده، \hat{Q}_{ni} مقدار نرمال‌شده دبی مدل‌شده و σ_ε انحراف‌معیار باقیمانده‌های مدل است. علاوه بر آن ضریب همبستگی⁸ (R)، میانگین قدرمطلق خطای نسبی⁹ (MARE)، ریشه میانگین مربعات خطا¹⁰ (RMSE)، میانگین خطا¹¹ (BIAS) و شاخص پراکندگی¹² (SI) بین دبی‌های مشاهده‌شده و مدل‌شده به‌صورت زیر محاسبه گردید:

$$R = \frac{\sum_{i=1}^n (Q_i - \bar{Q}_i)(\hat{Q}_i - \bar{\hat{Q}}_i)}{\sqrt{\sum_{i=1}^n (Q_i - \bar{Q}_i)^2 \sum_{i=1}^n (\hat{Q}_i - \bar{\hat{Q}}_i)^2}} \quad [14]$$

$$MARE = \frac{\sum_{i=1}^n \left| \frac{Q_i - \hat{Q}_i}{Q_i} \right|}{n} \quad [15]$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (Q_i - \hat{Q}_i)^2}{n}} \quad [16]$$

$$BIAS = \frac{\sum_{i=1}^n (Q_i - \hat{Q}_i)}{n} \quad [17]$$

$$SI = \frac{RMSE}{\bar{Q}_i} \quad [18]$$

که در آن Q_i دبی مشاهده‌شده، \hat{Q}_i دبی محاسبه‌شده، \bar{Q}_i میانگین دبی‌های مشاهده‌شده و $\bar{\hat{Q}}_i$ میانگین دبی‌های محاسبه‌شده است. برای تعیین توانایی‌های مدل‌ها در استفاده از نتایج آن‌ها در کوتاه‌مدت و درازمدت خطای نسبی (E_i) و میانگین تجمعی خطای نسبی (F_i) به‌صورت زیر برای داده‌های اعتبارسنجی محاسبه شد:

$$E_i = \frac{|Q_i - \hat{Q}_i|}{Q_i} \quad [19]$$

لایه چهارم (گره‌های تالی): در این لایه مقادیر پارامترهای p ، q و r بهینه می‌شوند. در واقع همه گره‌ها در این لایه با یک تابع گره به‌صورت زیر، سازگار می‌شوند:

$$O_i^4 = \bar{w}_i(p_i Q_{i-1} + q_i Q_{i-2} + r_i) \quad i=1,2 \quad [10]$$

که p_i ، q_i و r_i مجموع پارامترها و \bar{w} خروجی این لایه است.

لایه پنجم (گره خروجی): در این لایه همه خروجی‌ها به‌عنوان مجموع همه سیگنال‌های ورودی به‌صورت زیر، محاسبه می‌شود:

$$O_i^5 = \bar{a} \bar{w}_i f_i = \frac{\bar{a} w_i f_i}{\bar{a} w_i} \quad [11]$$

در این پژوهش برای آموزش ANFIS از دو الگوریتم پسانتشار خطا¹ و الگوریتم ترکیبی² که آمیخته‌ای از دو روش کاهش گرادیان³ و کمینه مربعات⁴ است، استفاده گردید.

معیارهای ارزیابی

معیار میانگین مربعات خطا⁵ (MSE) و معیار اطلاعات آکائیکه‌ی اصلاح شده⁶ (AICc) (آکائیکه 1987) برای تعیین بهترین مدل ARIMA بر اساس اصل امساک⁷ (مدلی با کمترین خطا و درعین‌حال کمترین تعداد پارامتر) به‌صورت زیر استفاده شد:

$$MSE = \frac{\sum_{i=1}^n (Q_{ni} - \hat{Q}_{ni})^2}{n - k - 1} \quad [12]$$

$$AICc = n \cdot \ln(\sigma_\varepsilon^2) + \frac{2kn}{n - k - 1} \quad [13]$$

¹ Back-Propagation algorithm (BP)

² Hybrid algorithm

³ Gradient descent rule

⁴ Least Squares Estimate (LSE)

⁵ Mean Squared Error (MSE)

⁶ Corrected Akaike's Information Criterion (AICc)

⁷ Principle of parsimony

⁸ Correlation coefficient (R)

⁹ Mean absolute relative error (MARE)

¹⁰ Root mean squared error (RMSE)

¹¹ Mean errors (BIAS)

¹² Scatter index (SI)

$$F_i = \frac{\sum_{j=1}^i a_j E_j}{i} \quad [20]$$

با توجه به شکل 1 و مطالب گفته شده در بخش مواد و روش‌ها، در مدل ARIMA با اعمال یکبار تفاضل‌گیری فصلی سری زمانی ایستا خواهد شد. از بین تمام مدل‌های ARIMA فقط 13 مورد دارای پارامترهای معنی‌دار و باقی‌مانده‌های مستقل بودند. جدول 2 نتایج معیارهای ارزیابی این مدل‌ها را نشان می‌دهد. در این جدول چهار مدل اول دارای بهترین نتایج در دوره‌های واسنجی و اعتبارسنجی می‌باشند. عملکرد این مدل‌ها بسیار شبیه به هم است. بر اساس معیارهای اصل امساک مدل $ARIMA(1,0,2)(0,1,1)$ به-عنوان مدل برتر انتخاب می‌گردد. جدول 3 نتایج آزمون‌های تی-استیودنت و باکس-پیرس را نشان می‌دهد. از آنجایی که P_τ برای هر پارامتر کمتر و P_Q برای هر تأخیر زمانی بیشتر از 5% است، این مدل دارای پارامترهای معنی‌دار و باقی‌مانده‌های مستقل است.

نتایج و بحث

تعیین بهترین مدل‌ها

برای مدل‌سازی سری‌های زمانی شرایط نرمال و ایستا بودن سری باید بررسی شود. برای بررسی نرمال بودن سری زمانی نتایج آزمون جارکو-برا قبل و بعد از اعمال تبدیل باکس-کاکس در جدول 1 ارائه شده است. مشاهده می‌شود که مقدار احتمال متناظر با آماره آزمون در توزیع مربع کای از سطح اعتماد 5% بیشتر است. این امر بیانگر نرمال شدن سری بعد از تبدیل است.

جدول 1- نتایج آزمون نرمال بودن جارکو-برا.

احتمال متناظر (%)	مقدار آماره	قبل از تبدیل	بعد از تبدیل
0/00	464/93	قبل از تبدیل	بعد از تبدیل
15/26	3/76	قبل از تبدیل	بعد از تبدیل

جدول 2- نتایج شاخص‌های آماری مختلف برای مدل‌های ARIMA.

SI	اعتبارسنجی			R	SI	واسنجی			R	اصل امساک		مدل‌های ARIMA
	BIAS	RMSE	MARE			BIAS	RMSE	MARE		MSE	AICc	
0/689	-1/541	1/272	0/398	0/762	0/896	-1/871	1/377	0/38	0/826	0/224	-356/5	(1.0.2)(0.1.1) ₁₂
0/687	-1/551	1/269	0/393	0/762	0/893	-1/88	1/372	0/382	0/827	0/225	-355/2	(2.0.1)(0.1.1) ₁₂
0/691	-1/536	1/277	0/399	0/759	0/893	-1/862	1/372	0/381	0/827	0/227	-354	(1.0.1)(0.1.1) ₁₂
0/703	-1/536	1/299	0/388	0/751	0/888	-1/859	1/364	0/391	0/83	0/229	-353	(1.0.0)(0.1.1) ₁₂
0/739	-1/633	1/365	0/443	0/72	0/938	-1/904	1/441	0/395	0/81	0/241	-334	(1.0.2)(5.1.0) ₁₂
0/756	-1/615	1/396	0/431	0/709	0/924	-1/906	1/42	0/404	0/816	0/246	-330/2	(1.0.1)(5.1.0) ₁₂
0/747	-1/609	1/379	0/433	0/716	0/89	-1/897	1/367	0/414	0/829	0/249	-328/8	(1.0.0)(5.1.0) ₁₂
0/716	-1/64	1/322	0/44	0/742	0/921	-1/899	1/415	0/403	0/815	0/257	-319/4	(1.0.2)(4.1.0) ₁₂
0/732	-1/633	1/352	0/435	0/733	0/903	-1/902	1/387	0/415	0/823	0/261	-317/9	(1.0.0)(4.1.0) ₁₂
0/703	-1/651	1/299	0/5	0/748	0/949	-1/906	1/457	0/405	0/803	0/267	-311/3	(1.0.2)(3.1.0) ₁₂
0/717	-1/638	1/325	0/493	0/738	0/925	-1/916	1/421	0/423	0/813	0/272	-309/5	(1.0.0)(3.1.0) ₁₂
0/749	-1/657	1/384	0/466	0/721	0/916	-1/932	1/407	0/432	0/819	0/286	-298/5	(1.0.0)(2.1.0) ₁₂
0/786	-1/662	1/452	0/569	0/685	0/9	-1/947	1/382	0/47	0/824	0/314	-277/2	(1.0.0)(1.1.0) ₁₂

جدول 3- نتایج آزمون‌های استقلال و معنی‌داری پارامترهای مدل ARIMA.

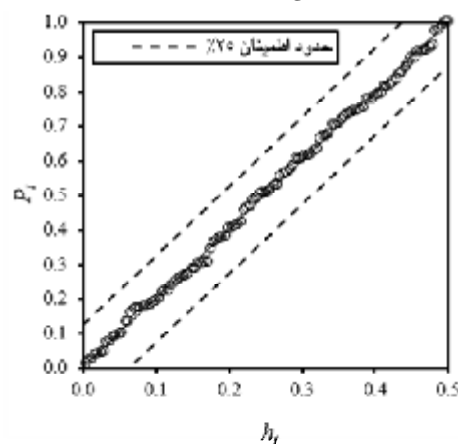
آزمون باکس-پییرس				آزمون تی-استیودنت				نام پارامتر
48	36	24	12	تأخیر زمانی	θ_1	θ_2	θ_3	φ_1
41/2	27/3	19/5	2/5	Q^*	0/92	0/15	0/195	0/9
43	31	19	7	DF	20/76	1/98	2/46	21/71
%55	%65/7	%42/4	%93	P_{Q^*}	%0	%4/8	%1/4	%0
								P_r

نتایج، مدل 6 که شامل ورودی‌های Q_{t-1} ، Q_{t-2} و Q_{t-12} هست در حالت ANFIS-GP-Hybrid دارای بهترین نتایج در دوره واسنجی و اعتبارسنجی هست. در نتیجه این مدل به‌عنوان بهترین مدل انتخاب شد. برای این مدل مقادیر به‌دست‌آمده برای شاخص‌های R ، MARE، RMSE، BIAS و SI به‌ترتیب برابر 0/768، 1/016، 1/551، 0/037، 0/838 هست.

توانایی مدل‌ها در مدل‌سازی دبی ماهانه

پس از انتخاب بهترین مدل از هر دو روش تصادفی و نروفازی که به‌ترتیب مدل‌های ANFIS-GP-Hybrid و $ARIMA(1,0,2)(0,1,1)_{12}$ می‌باشند، عملکرد این دو مدل در مدل‌سازی دبی ماهانه با یکدیگر مقایسه شد. شکل 3 نشان می‌دهد که هر دو مدل ANFIS و $ARIMA$ در دوره اعتبارسنجی نحوه تغییرات دبی نسبت به زمان را به‌خوبی شناسایی کرده‌اند به‌طوری‌که با افزایش (کاهش) مقدار واقعی دبی، مقادیر مدل‌سازی شده نیز افزایش (کاهش) می‌یابد. بنابراین این دو مدل دارای توانایی شناسایی مناسب ماهیت این پدیده‌ی هیدرولوژیکی هستند. بیشترین خطای مربوط به هر دو مدل در دبی‌های پیک مشاهده می‌شود که علت آن تغییرات فصلی بسیار شدید دبی ماهانه است. به‌عنوان مثال از سپتامبر 1994 تا دسامبر همان سال یعنی با گذشت فقط 3 ماه مقدار دبی بیش از 78 برابر شده است.

شکل 2 دوره‌نگار تجمعی مربوط به باقیمانده‌های مدل برتر ARIMA را نشان می‌دهد. همان‌گونه که از شکل پیداست نقاط به‌دست‌آمده داخل حدود اطمینان نشان داده‌شده در سطح اعتماد 95% هست. این موضوع بیانگر حذف کامل جزء تناوب از باقی‌مانده‌ها است که دلیل دیگری بر مستقل بودنشان است. زیرا جزء تناوب یک جزء قطعی (وابسته) است و اگر از باقی‌مانده‌ها حذف شود، سری باقی‌مانده‌ها مستقل خواهد بود.



شکل 2- دوره‌نگار تجمعی باقیمانده‌های مدل ARIMA.

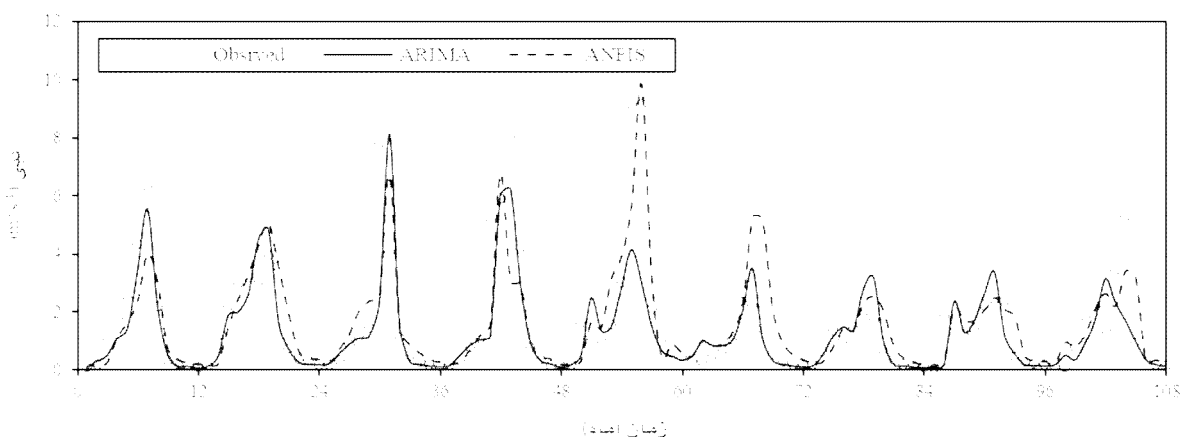
معیارهای ارزیابی عملکرد مدل ANFIS در مدل‌سازی دبی برای هر 14 نوع ترکیب ورودی مختلف محاسبه شد. در هر یک از مدل‌ها عملکرد دو الگوریتم پس‌انتشار خطا (BP) و الگوریتم ترکیبی (Hybrid) در یادگیری ANFIS در دوره واسنجی مورد بررسی قرار گرفت. در هر یک از دو الگوریتم یادگیری نیز از دو روش شبکه‌بندی (GP) و خوشه‌بندی تفریقی (SC) برای تولید سامانه‌ی استنتاج فازی استفاده گردید. با توجه به

ANFIS بیشتر است. نمودار شاخص F نیز نشان می‌دهد که مدل ARIMA در تمامی طول دوره اعتبارسنجی دارای مقادیر کمتری نسبت به مدل ANFIS است. نمودار شاخص F نشان می‌دهد که خطای مدل -ساز برای ماه اول در مدل ARIMA کمترین مقدار را دارد. مقدار این شاخص برای مدل ANFIS برای ماه اول 4/5 برابر مدل ARIMA است. یعنی مدل ARIMA نسبت به ANFIS در مدل‌سازی کوتاه‌مدت دبی ماهانه دقیق‌تر است. با گذشت زمان نمودار شاخص F برای مدل ARIMA نسبت به زمان تقریباً افقی شده است. به عبارت دیگر به نوعی رکود یا ایستایی رسیده است. در حالی که این نمودار برای مدل ANFIS صعودی است و با گذشت زمان افزایش می‌یابد. این موضوع بیانگر این است که با افزایش زمان استفاده از نتایج مدل‌ها در بلندمدت، خطای مدل ARIMA تقریباً ثابت می‌ماند. اما خطای مدل ANFIS افزایش خواهد یافت. در نتیجه مدل ARIMA نسبت به ANFIS از نظر استفاده از نتایج آن در درازمدت دارای دقت بیشتری در مدل‌سازی است.

شکل 4 دبی‌های مشاهده شده و مدل شده توسط مدل‌های ARIMA و ANFIS را در دوره واسنجی و اعتبارسنجی نشان می‌دهد. مشاهده می‌شود که نتایج به دست آمده برای دبی‌های با مقادیر کم نسبت به دبی‌های پیک در هر دو دوره به خط بهترین برازش نزدیک‌تر است. مدل ARIMA در مقایسه با ANFIS دبی‌های با مقادیر کم را دقیق‌تر برآورد کرده است. دقت هر دو مدل در مواقع سیلابی پایین می‌آید. هر دو مدل برای تمامی مقادیر دبی‌ها را هم بزرگ‌تر و هم کوچک‌تر از مقادیر واقعی تخمین زده‌اند که دبی‌های بزرگ‌تر، خطای کمتری نسبت به دبی‌های کوچک‌تر دارند.

استفاده از نتایج مدل‌ها در کوتاه‌مدت و درازمدت

شکل 5 نمودار تغییرات خطای نسبی (E) و میانگین تجمعی خطای نسبی (F) را نسبت به زمان نشان می‌دهد. با توجه به این شکل مشاهده می‌شود که مقادیر شاخص E برای مدل ANFIS بیشتر از مدل ARIMA است. همچنین نقاط اوج این نمودار برای مدل



شکل 3- دبی‌های مشاهده شده و مدل شده توسط مدل‌های ARIMA و ANFIS در دوره اعتبارسنجی.

تجمع داده‌ها و تغییرات یکنواخت در این محدوده دانست. دوم اینکه دقت این مدل در استفاده از نتایج‌اش در کوتاه‌مدت و بلندمدت بیشتر بود. مدل‌سازی کوتاه-مدت در مدیریت بهره‌برداری از سازه‌های آبی و

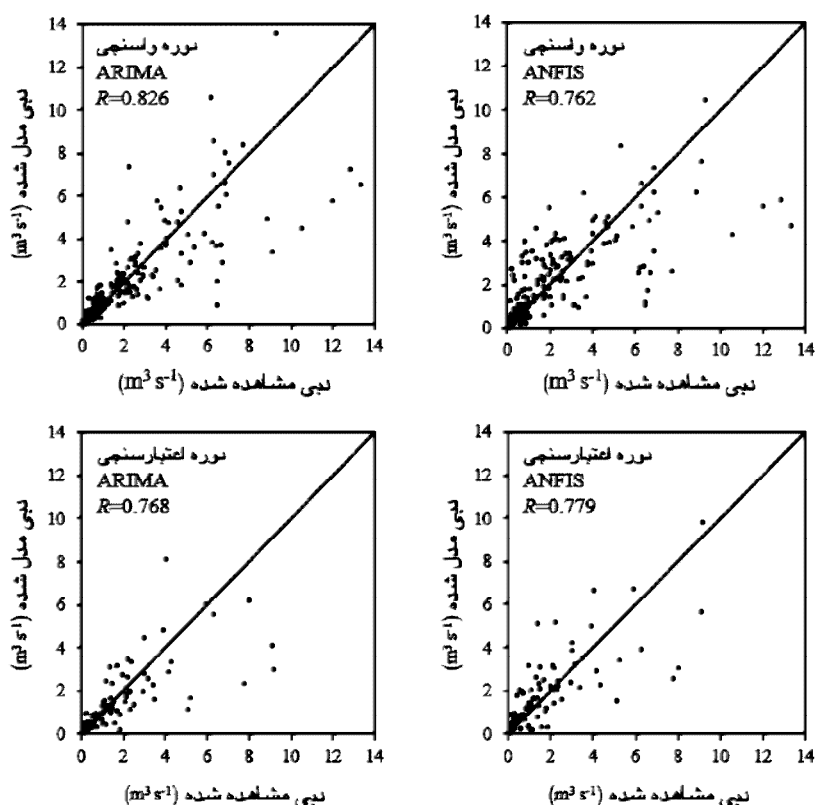
مشاهده شد که دقت هر دو مدل ARIMA و ANFIS تقریباً مشابه است؛ اما مدل ARIMA دارای دو مزیت جالب بود. اول اینکه برای مدل‌سازی دبی‌های با مقادیر کم بسیار دقیق‌تر بود که علت آن را می‌توان

غیریکنواختی زیاد است مانند دبی‌های ماهانه نتایج بهتری نسبت به مدل‌های پیچیده نروفازی دارند.

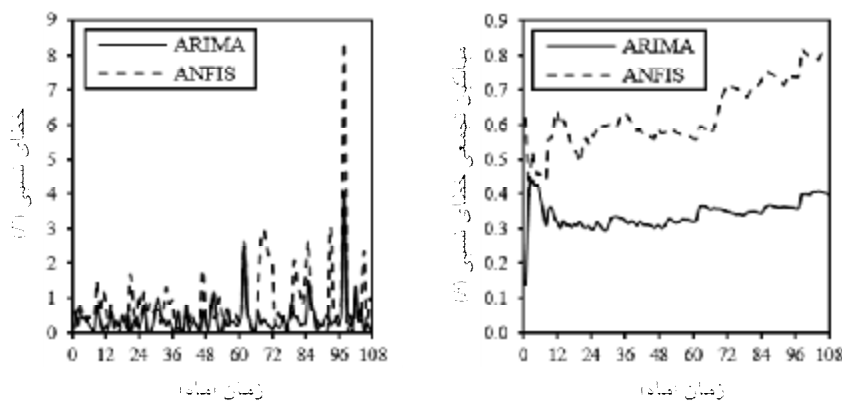
نتیجه‌گیری کلی

به‌منظور بررسی تأثیر تئوری‌های آمار و احتمالات و هوش مصنوعی در پیش‌بینی دبی ماهانه، مدل‌های ARIMA و ANFIS به‌کاربرده شد. هر دو مدل روند تغییرات دبی ماهانه نسبت به زمان را به‌خوبی شناسایی کردند. این مدل‌ها در دبی‌های پیک با افزایش خطا همراه بودند. برخلاف تصور که مدل‌های مبتنی بر هوش محاسباتی را دارای برتری قاطع نسبت به مدل‌های پایه می‌دانند، در این پژوهش هر دو مدل دارای نتایج تقریباً مشابهی بودند.

مدیریت توزیع آب مهم است و پیش‌بینی بلندمدت در طراحی سازه‌های آبی مثل سد کاربرد دارد. خلقی و همکاران (1385) در پیش‌بینی سری زمانی دبی‌های حداقل ماهانه رودخانه نیز مدل ARIMA را برتر از ANFIS معرفی کردند و علت آن را توانایی مدل‌های استوکستیک در مدل‌سازی بخش تصادفی بیان کردند. اما فیرات (2008) در مدل‌سازی دبی روزانه و نواز و همکاران (2015) در مدل‌سازی سطح رودخانه مدل ANFIS را برتر از مدل استوکستیک استفاده‌شده در پژوهش خود معرفی کردند. باید دقت کرد که دبی روزانه و سطح رودخانه تغییرات ملایمی در زمان دارند ولی دبی‌های ماهانه تغییرات فصلی شدید دارند. در نتیجه مدل‌های پایه استوکستیک در مواقعی که



شکل 4- دبی‌های مشاهده‌شده و مدل‌شده توسط مدل‌های ARIMA و ANFIS.



شکل 5- تغییرات خطای نسبی و میانگین تجمعی خطای نسبی در دوره اعتبارسنجی.

موضوع اهمیت در نظر گرفتن جزء تصادفی در مدل-سازی را به خصوص در مواقعی که با تغییرات فصلی شدید همراه هستیم به خوبی نشان می‌دهد.

حتی مدل ARIMA در مدل‌سازی دبی‌های با مقادیر کم و در مدل‌سازی به‌گونه‌ای که استفاده از نتایج آن در کوتاه‌مدت و درازمدت دارای دقت بیشتری باشد، عملکرد بسیار بهتری نسبت به ANFIS داشت. این

منابع مورد استفاده

- احمدی ف، آیشم س، خلیلی ک و بهمنش ج، 1394. کاربرد سیستم‌های استنتاج فازی-عصبی تطبیقی و ماشین بردار پشتیبان برای برآورد تبخیر و تعرق مرجع ماهانه شمال غرب کشور. نشریه پژوهش آب در کشاورزی، جلد 29، شماره 2، صفحه‌های 235 تا 247.
- احمدی ف، دین پژوهی، فاخری فردا و خلیلی ک، 1393. مقایسه مدل‌های خطی و غیرخطی سری زمانی در پیش‌بینی جریان رودخانه (مطالعه موردی: رودخانه باراندوزچای ارومیه). مجله علوم و مهندسی آبیاری، دوره 37، شماره 1، صفحه‌های 93 تا 105.
- احمدی ف، دین پژوهی، فاخری فردا، خلیلی ک و دربندی ص، 1394. مقایسه مدل‌های غیرخطی سری زمانی و برنامه‌ریزی ژنتیک در پیش‌بینی جریان روزانه رودخانه‌ها (مطالعه موردی: رودخانه باراندوزچای). مجله پژوهش‌های حفاظت آب و خاک، جلد 22، شماره 1، صفحه‌های 151 تا 169.
- خلقی م، اشرف‌زاده ا و مالمریم م، 1385. پیش‌بینی کم‌آبی ماهانه با استفاده از یک مدل استوکستیک و سیستم استنتاج فازی مبتنی بر شبکه تطبیقی. مجله تحقیقات منابع آب ایران، سال 5، شماره 2، صفحه‌های 16 تا 26.
- ناوه ه، خلیلی ک، اعلی م و بهمنش ج، 1391. پیش‌بینی جریان رودخانه با استفاده از مدل غیرخطی سری زمانی دو-خطی (مطالعه موردی رودخانه‌های باراندوزچای و شهرچای). نشریه علمی پژوهشی آب و خاک (علوم و صنایع کشاورزی)، جلد 26، شماره 5، صفحه‌های 1299 تا 1307.
- Akaike H, 1987. Factor analysis and AIC. *Psychometrika* 52(3): 317-32.
- Box GE and Cox DR, 1964. An analysis of transformations. *Journal of the Royal Statistical Society. Series B (Methodological)*: 211-52.
- Box GE and Pierce DA, 1970. Distribution of residual autocorrelations in autoregressive-integrated moving average time series models. *Journal of the American Statistical Association* 65: 1509-26.
- Cryer JD and Chan K-S, 2008. *Time Series Analysis With Applications in R*, Second ed. Springer, New York (ISBN: 0387759581, p. 491).
- Ebtehaj I and Bonakdari H, 2014. Performance Evaluation of Adaptive Neural Fuzzy Inference System for Sediment Transport in Sewers. *Water Resources Management* 28: 4765-79.

- Firat M, 2008. Comparison of artificial intelligence techniques for river flow forecasting. *Hydrology and Earth System Sciences* 12: 123-39.
- Hipel KW and McLeod AI, 1994. *Time Series Modelling of Water Resources and Environmental Systems*. Elsevier, Amsterdam.
- Hurst HE, Black RP and Simaika Y, 1965. *Long-Term Storage: an Experimental Study*. Constable, London.
- Jarque CM and Bera AK, 1980. Efficient tests for normality, homoscedasticity and serial independence of regression residuals. *Economics Letters* 6: 255-59.
- Khatibi R, Ghorbani MA, Naghipour L, Jothiprakash V, Fathima TA and Fazelifard MH, 2014. Inter-comparison of time series models of lake levels predicted by several modeling strategies. *Journal of Hydrology* 511: 530-545.
- Kisi O and Cigizoglu HK, 2007. Comparison of different ANN techniques in river flow prediction. *Civil Engineering and Environmental Systems* 24(3): 211-231.
- Kisi O, Dailr AH, Cimen M and Shiri J, 2012. Suspended sediment modeling using genetic programming and soft computing techniques. *Journal of Hydrology* 450: 48-58.
- McLeod A, 1978. On the distribution of residual autocorrelations in Box-Jenkins models. *Journal of the Royal Statistical Society. Series B (Methodological)*: 296-302.
- Nawaz N, Harun S and Talei A, 2015. Application of adaptive network-based fuzzy inference system (ANFIS) for river stage prediction in a tropical catchment. 735: 195-199. *Applied mechanics and materials*. Trans Tech Publ, Switzerland.
- Salas JD, Delleur JW, Yevjevich V and Lane WL, 1980. *Applied Modeling of Hydrologic Time Series*. Water Resources Publication, Colorado.
- Valipour M, 2012. Ability of Box-Jenkins models to estimate of reference potential evapotranspiration a case study, Mehrabad synoptic station, Tehran, Iran. *Journal of Agriculture and Veterinary Science* 1(5): 1-11.
- Valipour M, 2015. Long-term runoff study using SARIMA and ARIMA models in the United States. *Meteorological Applications* 22(3): 592-598.