

Technical Note

Evaluation of the Efficiency of Conventional and Computerized Methods for Reconstruction of Monthly Flow Time Series in Hydrometric Stations

H. Nozari^{1*} and F. Tavakoli²

Abstract

The lack of complete data should not hinder the study of the hydrological condition and the long-term forecasts for performing hydro-projects in one region. Various researchers have used different methods such as Ratio Analysis, Fragment, and Thomas-fiering for the reconstruction of incomplete flow data in hydrometric stations. So, in this study, the accuracy of these methods and computerized methods such as, artificial neural network, hybrid wavelet-neural network, and support vector machine have been investigated and compared. The results showed that the computerized methods have the higher accuracy than the other three methods. Comparison amongst the computerized methods showed that the artificial neural network method ($R^2 = 0.98$, RMSE = 6.18, SE = 0.476), the support vector machine method ($R^2 = 0.902$, RMSE = 6.074, SE = 0.486), and the hybrid wavelet-neural method ($R^2 = 0.889$, RMSE = 6.96, SE = 0.54) ranked first to third, respectively. Although, these three methods have not significant differences in results, but the support vector machine constructed the data in the less time and with more ease and hence had an advantage in comparison with the other methods.

ارزیابی کارایی روش‌های مرسوم و رایانه‌ای در بازسازی سری زمانی دبی ماهانه ایستگاه‌های هیدرومتری

حامد نوذری^{۱*} و فاطمه توکلی^۲

چکیده

عدم وجود آمار و اطلاعات کامل، نمی‌تواند مجوزی برای عدم مطالعه شرایط هیدرولوژیکی یک منطقه و پیش‌بینی‌های درازمدت برای انجام یک پروژه آبی باشد. بنابراین، پژوهشگران مختلف روش‌هایی از قبیل آنالیز نسبت‌ها، فرگمنت و توماس فیرینگ را برای بازسازی داده‌های ناقص دبی در ایستگاه‌های هیدرومتری به کار برده‌اند؛ لذا در این پژوهش دقت روش‌های مذکور با روش‌های رایانه‌ای از قبیل شبکه عصبی مصنوعی، هیرید عصبی-موجکی و ماشین بردار پشتیبان مورد مقایسه و بررسی قرار گرفته است. نتایج نشان می‌دهند که روش‌های رایانه‌ای نسبت به سه روش دیگر از دقت بالاتری برخوردار هستند. مقایسه نتایج روش‌های رایانه‌ای بیانگر آن است که شبکه عصبی مصنوعی ($R^2 = 0.98$, RMSE = 6.18, SE = 0.476) و ماشین بردار پشتیبان ($R^2 = 0.902$, RMSE = 6.074, SE = 0.486) و هیرید عصبی-موجکی ($R^2 = 0.889$, RMSE = 6.96, SE = 0.54) به ترتیب در رتبه‌های اول تا سوم قرار دارند. اگرچه سه روش شبکه عصبی مصنوعی، ماشین بردار پشتیبان و هیرید عصبی-موجکی تفاوت معناداری در مقایسه با یکدیگر ندارند؛ اما روش ماشین بردار پشتیبان با سهولت بیشتر و در زمان کمتری قادر به بازسازی داده‌ها بوده و از این جهت نسبت به سایر روش‌ها ارجحیت دارد.

کلمات کلیدی: بازسازی، حوضه آبریز دریاچه ارومیه، داده گمشده، دبی ماهانه، ماشین بردار پشتیبان، هیرید عصبی موجکی.

تاریخ دریافت مقاله: ۹۵/۷/۱۸

تاریخ پذیرش مقاله: ۹۶/۲/۶

Keywords: Hybrid Wavelet-Neural Network, Missing Data, Monthly Flow, Reconstruction, Support Vector Machine, Urmia Lake Basin.

Received: October 9, 2016

Accepted: April 26, 2017

1- Assistant Professor, Water Science and Engineering Department, Faculty of Agriculture, Bu-Ali Sina University, Hamedan. Email: hanozari@yahoo.com
2- M.Sc. Student in Water Resources Engineering, Faculty of Agriculture, Bu-Ali Sina University, Hamedan.

*- Corresponding Author

۱- استادیار گروه علوم و مهندسی آب، دانشکده کشاورزی، دانشگاه بوعلی سینا، همدان
۲- دانشجوی کارشناسی ارشد مهندسی منابع آب، دانشکده کشاورزی، دانشگاه بوعلی سینا، همدان

*- نویسنده مستول (Discussion) در مورد این مقاله تا پایان خرداد ۱۳۹۷ امکانپذیر است.

۱- مقدمه

پارامتریک فرگمنت، توماس فیرینگ و Valencia-Schaak در تولید داده‌های ماهیانه مصنوعی را مورد مقایسه و ارزیابی قرار دادند. در تحقیق حاضر به منظور معرفی مدلی مناسب چهت بازسازی سری زمانی دبی ماهانه ایستگاه‌های هیدرومتری، سعی شد از روش‌های مرسم و نویزی که مبتنی بر آمار و اطلاعات پیشین خود ایستگاه و یا ایستگاه‌های مجاور می‌باشد مانند روش توماس فیرینگ، آنالیز نسبت‌ها، فرگمنت، شبکه عصبی مصنوعی، هیبرید عصبی-موجکی و ماشین بردار پشتیبان، استفاده شود و نتایج هر یک مورد ارزیابی قرار گرفته و بهترین روش بازسازی آبدی رودخانه، معروفی گردد.

۲- معرفی منطقه مورد مطالعه

حوضه آبریز دریاچه ارومیه در شمال غرب ایران به وسیله بخش شمالی کوههای زاگرس و دامنه جنوبی کوه سبلان و نیز دامنه‌های شمالی، غربی و جنوبی کوه سهند احاطه شده است. مساحت حوضه مذکور ۵۱۸۰۱ کیلومتر مربع بوده که حدود ۶۴/۶ درصد آن را مناطق کوهستانی، ۲۱/۱ درصد آن را دشت‌ها و کوهپایه‌ها و ۱۴/۳ درصد آن را دریاچه ارومیه تشکیل می‌دهد. در تحقیق حاضر از آمار و اطلاعات ماهانه ۱۵ سال مربوط به ۹ ایستگاه هیدرومتری واقع در این حوضه با نام‌های نقده، کوتر، میاندوآب، قبقلو، گرزال، بیطاس، پل بهراملو، داشنید بوکان و پل سردشت (بریسو) استفاده شد.

۳- مواد و روش‌ها

۳-۱- شبکه عصبی مصنوعی

این روش یک سیستم هوشمند است که با پردازش روی داده‌های تجربی و کشف قوانین حاکم بر آنها اقدام به پیش‌بینی می‌نماید (Sattari et al., 2014). شبکه مذکور مشتمل از لایه‌های ورودی، مخفی و خروجی بوده و با استفاده از پردازشگرهایی به نام نرون، نگاشتی از فضای ورودی به فضای مطلوب ارائه می‌دهد (Ghafari and Vafakhah, 2013).

۳-۲- تئوری موجک

تبديل موجک یکی از تبدیل‌های ریاضی کارآمد در زمینه آنالیز سیگنال و استخراج اطلاعاتی است که دستیابی به آن‌ها از طریق سیگنال‌های خام به راحتی امکان پذیر نیست (Roushangar et al., 2015). یکی از مهمترین قابلیتها و کاربردهای آنالیز موجک، کمک به پیش‌بینی سری‌های زمانی با استفاده از تجزیه و تبدیل سیگنال‌ها به امواج پایه و اختلالات در مراتب گوناگون است.

اولین گام در اجرای طرح‌ها و پروژه‌های مربوط به منابع آب داشتن آمار و اطلاعات صحیح و کافی می‌باشد؛ اما در بیشتر ایستگاه‌های هیدرومتری، خلاهای آماری گسترده‌ای به چشم می‌خورد. لذا قبل از به کارگیری این اطلاعات باید داده‌ها مورد بازنگری قرار گرفته و تکمیل شوند. در این راستا دستیابی به روشنی صحیح به منظور بازسازی و تکمیل آمار بسیار ضروری بوده و برآورد دقیق آن، باعث افزایش دقت در برآورد پارامترهای هیدرولوژیکی و در نهایت باعث کاهش خسارات ناشی از اجرای طرح‌ها می‌شود (Naghadi et al., 2010; matinzadeh et al., 2011). با مطالعه پژوهش‌های صورت گرفته پیشین می‌توان گفت، استفاده از روش‌های سنتی در بازسازی و تکمیل سری‌های زمانی هیدرولوژیکی از گذشته مرسوم بوده است؛ اما در سال‌های اخیر استفاده از روش‌های رایانه‌ای نیز در این زمینه مورد توجه پژوهشگران قرار گرفته است. Coulibaly and Evora (2007) با استفاده از مدل شبکه عصبی و شبکه عصبی دینامیکی بازگشتی به تکمیل داده‌های روزانه بارش در ۱۵ ایستگاه هواشناسی در شمال شرق کانادا پرداختند. Naghadi et al. (2010) به منظور بازسازی آمار دبی ماهانه ایستگاه‌های هیدرومتری حوضه آبخیز کارون بزرگ از روش‌های شبکه عصبی مصنوعی، نسبت نرمال محور مختصات، رگرسیون خطی ساده و خودهمبستگی سری‌های زمانی استفاده کارایی روش رگرسیون خطی نمودند. Sadati Nejad et al. (2011) فازی در مقایسه با روش‌های رگرسیون خطی ساده، رگرسیون چند متغیره، روش محور مختصات و نسبت نرمال را در بازسازی دبی سالانه ایستگاه‌های هیدرومتری حوضه آبخیز کارون بزرگ مورد ارزیابی قرار دادند. Jalilian et al. (2012) از روش‌های رگرسیون تک متغیره، رگرسیون دو متغیره، لانگبین و توماس فیرینگ برای بازسازی دبی ماهانه، دبی سالانه، رسوب و کیفیت آب حوضه آبخیز سفیدرود استفاده کردند. Tfwala et al. (2013) داده‌های گمشده دبی ماهانه ایستگاه لی لین در جنوب تایوان را با استفاده از دو روش شبکه عصبی (MLP) و مدل CANFISM شبیه‌سازی نمودند. Nozari and Kadkhodahoseini (2014) داده‌های بارش ایستگاه باران‌سنگی سرابی با استفاده از آمار ایستگاه‌های مجاور آن از روش‌های شبکه عصبی مصنوعی و آزمون نسبت‌ها و تفاضل‌ها استفاده نمودند. Pourreza bilondi et al. (2015) سیلاب روزانه در حوضه رودخانه لیف آمریکا را با دو روش ماشین بردار پشتیبان و Hymod پیش‌بینی کردند. Montaseri and Heydari (2016) با استفاده از روش شبیه‌سازی مونت کارلو، عملکرد چهار تیپ از مدل غیرپارامتریک Bootstrap و مدل‌های

۴- بحث و نتایج

به منظور بررسی دقت روش‌های مذکور، آمار و اطلاعات ماهانه، ۱۱ سال از ایستگاه‌های مورد نظر به عنوان داده‌های مشاهده‌ای بدون نقص و آمار ۴ سال به عنوان داده‌های گمشده در نظر گرفته شد و این داده‌ها مورد بازسازی قرار گرفت. نتایج مربوط به محاسبه شاخص‌های آماری ضریب همبستگی، جذر میانگین مربعات خطأ و خطای استاندارد برای روش آنالیز نسبت‌ها، فرگمنت و توماس فیرینگ در جدول ۱ ارائه شده است. همانطور که ملاحظه می‌شود خطای استاندارد در روش آنالیز نسبت‌ها مابین ۰/۳۵ و ۰/۱۶، در روش فرگمنت مابین ۰/۳۶ و ۰/۱۸ و در روش توماس فیرینگ مابین ۰/۸۲ و ۰/۹۸ قرار دارد؛ لذا می‌توان گفت دقت روش آنالیز نسبت‌ها بالاتر از دو روش دیگر می‌باشد.

در این قسمت به منظور اجرای مدل شبکه عصبی مصنوعی از جعبه ابزار پردازش شبکه عصبی در نرم‌افزار MATLAB (2014) استفاده شد. این جعبه ابزار در حل مسائل از شبکه دو لایه پیشخور با الگوریتم پس انتشار خطأ، تابع انتقال خطی در لایه خروجی و تابع انتقال سیگموئید در لایه مخفی استفاده می‌نماید. جدول ۲ بهترین ساختارهای پیشنهادی شبکه در هریک از ایستگاه‌ها را نشان می‌دهد. همانطور که ملاحظه می‌شود در این روش خطای استاندارد مابین ۰/۱۱ و ۰/۸۰ قرار می‌گیرد که نسبت به سه روش قبل از دقت قابل قبولی برخوردار است.

اما به منظور اجرای مدل هیبرید عصبی موجکی، دو گام اصلی وجود دارد که عبارتند از، انتخاب موجک مادر و تعیین سطوح تجزیه، و با توجه به تأثیر مستقیم این دو عامل در نتایج شبیه‌سازی، لازم است انتخاب به دقت صورت پذیرد. در همین راستا و به منظور انتخاب موجک مادر از سه گروه موجک هار (HAAR)، سیملت (SYMLET) و دابچیز (DAUBECHIES) استفاده شد.

۳-۳- ماشین بردار پشتیبان

ماشین بردار پشتیبان یک سیستم یادگیری کارآمد است که از اصل استقرایی کمینه‌سازی خطای ساختاری استفاده می‌کند تا به یک جواب بهینه برسد. اولین کاربرد این روش در حل مسائل بارش و رواناب (Abdollah Pour Azad and Sattari, 2013) توسط دیباک و همکاران ارائه شده است. غیرخطی را حل می‌نماید که انتخاب کرنل مناسب وابسته به حجم داده‌های آموزشی می‌باشد (Ahmadi et al., 2015).

۴-۳- مدل توماس فیرینگ

مدل توماس فیرینگ یک مدل آماری برای پیش‌بینی پارامتر بر اساس زنجیره مارکوف است. در این روش هر داده مربوط به دبی ماهانه فقط به داده ماه قبل از خود وابسته بوده و با استفاده از رابطه زیر محاسبه می‌گردد (Jalilian et al., 2012)

$$q_{i+1} = q_{avj+1} + \left(\frac{s_{j+1}}{s_j} \right) r_{j,j+1} (q_j - q_{avj}) + \frac{z_i s_{j+1}}{\sqrt{1 - r_{j,j+1}^2}} \quad (1)$$

در این رابطه: q_{j+1} و q_j مقادیر دبی ماهانه به ترتیب در ماههای $j+1$ و j ؛ q_{avj+1} میانگین مقادیر دبی ماهانه به ترتیب برای ماههای $j+1$ ، j و s_{j+1} انحراف معیار برای ماههای $j+1$ ، j و $j-1$ ؛ $r_{j,j+1}$ ضریب همبستگی بین مقادیر دبی ماههای $j+1$ و j و z_i عدد تصادفی با توزیع نرمال (۰/۰) می‌باشد.

۵-۳- روش فرگمنت

این روش با استفاده از یک رویکرد صرفاً قطعی به تجزیه جریان سالانه به ماهانه می‌پردازد. در این روش، برای سال K ام، مقادیر جریان ماهانه مشاهده‌ای به جریان سالانه تقسیم می‌شوند و در نتیجه مجموعه‌ای از ۱۲ جریان استاندارد ماهانه تشکیل می‌شود (Silva and Portela, 2012).

Table 1- Reconstruction result by Ratio Analysis, Fragment, and Thomas-fiering methods

جدول ۱- نتایج بازسازی به روش آنالیز نسبت‌ها، فرگمنت و توماس فیرینگ

Station		Ratio Analysis			Fragment			Thomas-fiering		
Dependent	Independent	R ²	RMSE (m ³ /s)	SE	R ²	RMSE (m ³ /s)	SE	R ²	RMSE (m ³ /s)	SE
Ghabghabloo	Miandoab	0.69	13.75	0.98	0.65	13.11	0.94	0.53	13.53	0.97
Bitas	Kooter	0.64	1.6	0.89	0.5	1.86	1.04	0.56	3.56	1.98
Garjal	Barisoo	0.86	15.52	0.35	0.861	15.7	0.36	0.65	35.24	0.82
Miandoab	Dashboard bukan	0.49	18.6	1.16	0.47	18.9	1.18	0.36	27.95	1.72
Naghadeh	Pole bahramloo	0.87	6.9	0.7	0.73	8.15	0.82	0.6075	14.89	1.51

Table 2- Reconstruction results by artificial neural network method**جدول ۲- نتایج بازسازی به روش شبکه عصبی مصنوعی**

Station		Network structure		Statistical indicators		
Dependent	Independent	Number of Layers	Number of neurons	R ²	RMSE (m ³ /s)	SE
Ghabghabloo	Miandoab	2	1-4	0.815	11.2	0.80
bitas	Kooter	2	1-4	0.91	0.79	0.44
Garjal	Barisoo	2	1-4	0.988	4.9	0.11
Miandoab	Dashboard bukan	2	1-3	0.92	7.62	0.47
Naghadeh	Pole bahramloo	2	1-3	0.89	5.19	0.52

۵- نتیجه‌گیری

در این تحقیق دقت روش‌های آنالیز نسبت‌ها، فرگمنت، توماس فیرینگ، شبکه عصبی مصنوعی، هیبرید عصبی- موجکی و ماشین بردار پشتیبان در بازسازی داده‌های گم شده دبی ماهانه ایستگاه‌های هیدرومتری مورد بررسی قرار گرفته و نتایج زیر حاصل شد:

- در بین روش‌های آنالیز نسبت‌ها، توماس فیرینگ و فرگمنت، در مجموع روش آنالیز نسبت‌ها از توانایی بیشتری در بازسازی برخوردار است.

- به علت مطلق نبودن برتری یک روش در ایستگاه‌های مورد مطالعه و به منظور دستیابی به پاسخی مطمئن، میانگین پارامترهای آماری در ایستگاه‌های مختلف محاسبه شد. نتایج نشان داد، روش‌های شبکه عصبی مصنوعی، هیبرید عصبی- موجکی و ماشین بردار پشتیبان از دقت قابل ملاحظه‌ای در بازسازی داده‌های دبی ماهانه ایستگاه‌های هیدرومتری نسبت به روش‌های آنالیز نسبت‌ها، توماس فیرینگ و فرگمنت برخوردار هستند. نتایج این تحقیق با نتایج حاصل از تحقیقات سایر پژوهشگران مانند Naghadi et al. (2010) مطابقت دارد.

- روش‌های نوین شبکه عصبی مصنوعی، هیبرید عصبی- موجکی و ماشین بردار پشتیبان تفاوت معنی داری در مقایسه با یکدیگر ندارند.

برای تعیین سطح تجزیه مناسب، آموزش مدل با سطوح تجزیه ۲ الی ۱۰ صورت گرفت و نهایتاً موجک مادر Db4 با تعداد ۲ لایه و ۳ نرون به عنوان ساختار اصلی شبکه انتخاب شد. نتایج مربوط به محاسبه شاخص‌های آماری این روش در ایستگاه‌های مورد مطالعه در جدول ۳ نشان داده شده است. همانطور که ملاحظه می‌شود خطای استاندارد مابین ۰/۰۸ و ۰/۰۸۰ قرار دارد و بسیار نزدیک به روش شبکه عصبی مصنوعی می‌باشد.

از سوی دیگر مدل ماشین بردار پشتیبان در نرم‌افزار MATLAB (2014) کدنویسی و اجرا شد. برای انتخاب کرنل مناسب در هر ایستگاه، آموزش شبکه به وسیله چهار کرنل تابع پایه شعاعی (RBF)، چند جمله‌ای (POLY)، خطی (LIN) و تاثرانت هیپربولیک (MLP) صورت گرفت و در نهایت با توجه به شاخص‌های آماری میان دبی مشاهداتی و شبیه‌سازی شده توسط این مدل، کرنل RBF به عنوان مناسب‌ترین کرنل در هر ایستگاه انتخاب گردید. با توجه به جدول ۳ می‌توان گفت دقت این روش در مقایسه با روش‌های سنتی بازسازی به شدت افزایش یافته است ولی نسبت به روش‌های رایانه‌ای شبکه عصبی و موجک اختلاف قابل ملاحظه‌ای ندارد و خطای استاندارد آن مابین ۰/۰۸۵ و ۰/۰۸۱ می‌باشد.

Table 3- Reconstruction results by hybrid wavelet-neural network and support vector machine methods**جدول ۳- نتایج بازسازی به روش‌های هیبرید عصبی- موجکی و ماشین بردار پشتیبان**

Station		Wavelet-neural network			Support vector machine		
Dependent	Independent	R ²	RMSE (m ³ /s)	SE	R ²	RMSE (m ³ /s)	SE
Ghabghabloo	Miandoab	0.8337	11.24	0.8	0.79	11.91	0.85
bitas	Kooter	0.8672	1.1	0.61	0.9	0.87	0.48
Garjal	Barisoo	0.9838	6.9	0.16	0.99	4.7	0.1
Miandoab	Dashboard bukan	0.8277	11.19	0.7	0.92	8	0.5
Naghadeh	Pole bahramloo	0.9326	4.4	0.45	0.88	5.58	0.56

- the monthly streamflow and rainfall data. *Iran-Water Resources Research* 11(3):69-84 (In Persian)
- Naghadi R, Shayan Nejad M, Sadati Nejad S J (2010) Comparison of different methods for estimating of monthly discharge missing data in grand keroon river basin. *Journal of Watershed Management Research* 1(1):59-73 (In Persian)
- Nozari H, Kadkhodahoseini M (2014) Neural network and test performance ratios and differences in data reconstruction of the stations. In: 5th National Conference on Water Resources Management
- Pourreza Bilondi M, Khashei Siuki A, Sadeghi Tabas S (2015) Daily rainfall-runoff modeling with least square support vector machine(ls-svm). *Journal of Water and Soil Conservation* 21(6):293-304 (In Persian)
- Roushangar K, Zarghami M, Tarlani Azar M (2015) Forecasting daily urban water consumption using conjunctive evolutionary algorithm and wavelet transform analysis. *Journal of Water and Wastewater* 26(4):110-120 (In Persian)
- Sadati Nejad S J, Naghadi R, Shayan Nejad M (2011) Application of fuzzy linear regression for predicting annual discharge missing data in hydrometric station compared with other conventional method. *Journal of Water and Soil Conservation* 17(4):67-86 (In Persian)
- Sattari M T, Joudi A R, Nahrein F (2014) Monthly rainfall prediction using artificial neural networks and M5 model tree (case study: station of ahar). *Physical Geography Research Quarterly* 46(2):247-260 (In Persian)
- Silva AT, Portela M M (2012) Disaggregation modeling of monthly stream flows using a new approach of the method of fragments. *Hydrological Sciences Journal* 57(5):942-955
- Tfwala S, Min Wang Y, Chieh Lin Y (2013) Prediction of missing flow records using multilayer perceptron and coactive neurofuzzy inference system. *Scientific World Journal* 2013:1-7
- در بین روش‌های نوین بازسازی در مجموع شبکه عصبی مصنوعی، ماشین بردار پشتیبان و هیبرید عصبی - موجکی به ترتیب در رتبه‌های اول تا سوم قرار دارند.
- روش ماشین بردار پشتیبان علیرغم نتایج به دست آمده با سهولت بیشتر و کار و زمان کمتری بازسازی را انجام می‌دهد و نسبت به سایر روش‌ها ارجحیت دارد.
- ## - مراجع -
- Abdollah Pour Azad M, Sattari M T (2015) Forecasting daily river flow of ahar chay river using artificial neural networks (ANN) and comparison with adaptive neuro fuzzy inference system (ANFIS). *Journal of Water and Soil Conservation* 22(1):287-298 (In Persian)
- Ahmadi F, Radmanesh F, Mirabbasi Najaf Abadi R (2015) Comparison between genetic programming and support vector machine methods for daily River flow forecasting. *Journal of Water and Soil* 28(6):1162-1171 (In Persian)
- Coulibaly P, Evora N D (2007) Comparison of neural network methods for infilling missing daily weather records .*Journal of Hydrology* 341:27-41
- Ghafari GH, Vafakhah M (2013) Simulation of rainfall-runoff process using artificial neural network and adaptive neuro fuzzy interface system. *Journal of Watershed Management Research* 4(8):120-136 (In Persian)
- Jalilian H , Rostamizad GH , Arekhi S (2012) Applicability of various reconstruction methods of hydrometric data (case study: sefidroud basin). *Journal of Range and Watershed Management* 65(3):329-340 (In Persian)
- Matinzadeh M M , Fattahi R, Shahyan Nejad M, Abdollahi KH (2011) Reconstruction of annual maximum 24-h rainfall data using fuzzy regression in ch&b province. *Journal of Iran Water Research* 5(8):186-179 (In Persian)
- Montaseri M, Heydari J (2016) A comparison among the performance of the stochastic models in generating