

ترکیب بهینه متغیرها برای شبیه‌سازی رواناب در حوزه آبخیز امامه با استفاده از آزمون گاما

علیرضا شریفی^{1*}، یعقوب دین‌پژوه²، احمد فاختری‌فرد³ و علیرضا مقدم‌نیا⁴

تاریخ دریافت: 91/07/11 تاریخ پذیرش: 91/12/07

¹ دانشجوی سابق کارشناسی ارشد منابع آب، گروه مهندسی آب، دانشکده کشاورزی، دانشگاه تبریز

² دانشیار، گروه مهندسی آب، دانشکده کشاورزی، دانشگاه تبریز

³ استاد، گروه مهندسی آب، دانشکده کشاورزی، دانشگاه تبریز

⁴ دانشیار، گروه احیای مناطق خشک و کوهستانی، دانشکده منابع طبیعی، دانشگاه تهران

* مسئول مکاتبات، پست الکترونیکی: areza_sharifi@yahoo.com

چکیده

رواناب ناشی از بارش یک فرایند پیچیده و غیرخطی بوده و بنابراین، مدل‌سازی آن چندان آسان نیست. هدف این مطالعه کاربرد آزمون گاما برای انتخاب ترکیب بهینه متغیرهای ورودی در مدل‌سازی رواناب رودخانه حوزه آبخیز امامه می‌باشد. برای تعیین بهینه تعداد داده‌های مورد نیاز برای مدل‌سازی از آزمون M استفاده شد. داده‌های بارندگی $P(t)$ و رواناب $R(t)$ در مقیاس روزانه و در طول دوره آماری 1379-1388 استفاده شد. همچنین هشت متغیر ورودی شامل سری مربوط به جریان با تأخیر یک روزه $(R(t-1))$ ، دو روزه $(R(t-2))$ ، سه روزه $(R(t-3))$ و چهار روزه $(R(t-4))$ ، سری بارندگی روزانه بدون تأخیر زمانی $(P(t))$ و با تأخیرهای یک روزه $(P(t-1))$ ، دو روزه $(P(t-2))$ و سه روزه $(P(t-3))$ استفاده شد. مدل‌سازی جریان آب رودخانه با استفاده از تعداد نقاط بهینه متغیرهای منتخب با روش‌های شبکه عصبی مصنوعی و رگرسیون خطی محلی انجام شد. نتایج نشان داد که شش متغیر شامل $P(t)$ ، $P(t-1)$ ، $P(t-2)$ ، $P(t-3)$ و $R(t-1)$ و $R(t-2)$ بهینه ترکیب متغیرها در مدل‌سازی جریان رودخانه حوزه مذکور می‌باشند. همچنین با استفاده از خروجی آزمون M تعداد 1405 داده برای بخش آموزش مدل‌سازی مناسب تشخیص داده شد. نتایج حاکی از این واقعیت است که روش رگرسیون خطی محلی (LLR) در قسمت آموزش از دقت بالاتری نسبت به روش شبکه‌های عصبی مصنوعی برخوردار است، در حالیکه در مرحله تست مدل، روش شبکه عصبی از دقت بیشتری برخوردار بود. مقدار R^2 و RMSE روش LLR در بخش آموزش بترتیب معادل 0/96 و 1/7 بدست آمد.

واژه‌های کلیدی: آزمون گاما، حوزه آبخیز امامه، رگرسیون خطی محلی، شبکه عصبی مصنوعی

Optimum Combination of Variables for Runoff Simulation in Amameh Watershed using Gamma test

AR Sharifi¹*, Y Dinpashoh², A Fakheri-Fard³ and AR Moghaddamnia⁴

Received: 2 October 2012 Accepted: 25 February 2013

¹- Former M.Sc Student of Water Resources, Dept. of Water Engin., Faculty of Agric., Univ. of Tabriz, Iran

²- Assoc. Prof., Dept. of Water Engin., Faculty of Agric., Univ. of Tabriz, Iran

³- Prof., Dept. of Water Engin., Faculty of Agric., Univ. of Tabriz, Iran

⁴- Assoc. Prof., Dept. of Arid and Mountainous Regions Reclamation, Faculty of Natural Resources, Univ. of Tehran, Iran

* Corresponding Author Email: areza_sharifi@yahoo.com

Abstract

Runoff resulting from rainfall is a complex and non-linear process and, therefore, its modeling is not so easy. The aim of this study was the application of the Gamma test to select the optimal combination of input variables for runoff modeling in Amameh watershed. M-test was used to identify the optimal number of required data for modeling. Data of rainfall (P(t)) and runoff (R(t)) in daily time scale were used in the period of 2000 to 2009. Totally, eight input variables namely four variables of daily streamflow: lag-1 (R(t-1)), lag-2 (R(t-2)), lag-3 (R(t-3)) and lag-4 (R(t-4)) as well as four daily rainfall variables: without lag time (P(t)), lag-1 (P(t-1)), lag-2 (P(t-2)) and lag-3 (P(t-3)) were used. Streamflow modeling was performed based on the optimum number of the selected variables using the artificial neural network (ANN) and local linear regression (LLR) methods. The results showed that the six variables of P(t), P(t-1), P(t-2), P(t-3), R(t-1) and R(t-2) belonged to the optimum combination of variables in streamflow modeling of the mentioned watershed. Moreover, based on the M-test output, only 1405 points were found to be adequate for modeling in the training section. Results indicated the fact that the LLR method had greater accuracy in training process compared to ANN. However, the ANN had large amount of accuracy in the model testing process. In training section the R² and RMSE values were found to be equal to 0.96 and 1.7, respectively.

Keywords: Amameh watershed, Artificial neural network, Gamma test, Local linear regression

اقليمی، کمیت و کیفیت آب قابل دسترس برای توسعه اقتصادی و زیست محیطی در اکثر نقاط ایران کافی نمی‌باشد. در سال‌های گذشته وقوع سیلاب‌ها و

مقدمه

آب یکی از عوامل مهم در توسعه پایدار هر منطقه می‌باشد. به نظر می‌رسد با توجه به شرایط

مشاهداتی لازم است برای آموزش مدل در نظر گرفته شود. صادقی و همکاران (1383) نیز متغیرهای مرتبط و مشابه را استفاده کرده‌اند.

یکی از مراحل مهم و پیچیده برای مدل‌سازی غیر خطی، پیش پردازش داده‌های ورودی به منظور انتخاب ترکیبی مناسب از آنها در مدل می‌باشد. این کار سبب کاهش مراحل سعی و خطا و شناخت مهم‌ترین متغیرهای موثر بر پدیده مورد نظر در مدل‌سازی می‌گردد. برای این منظور تا به حال از روش‌های مختلف مانند روش تجزیه به مولفه‌های اصلی (ژانگ و همکاران 2006، ژانگ 2007، نوری و همکاران 2009)، پروکراتس (دین‌پژوه و همکاران 2004) و آزمون گاما (کورکورن و همکاران 2003، مقدم‌نیا و همکاران 2008) استفاده شده است. در مطالعه حاضر برای یافتن بهترین ترکیب از متغیرهای ورودی به مدل پیش‌بینی جریان آب رودخانه روش آزمون گاما انتخاب و مورد استفاده قرار گرفت. همچنین با استفاده از آزمون M می‌توان تعداد داده‌های مورد نیاز برای بخش آموزش مدل را هم تعیین کرد. آزمون گاما اولین بار توسط کانکار (1997) و سپس توسط استفانسن و همکاران (1997) معرفی شد. ایوانس (2001) و ایوانس و جونز (2002) از این روش برای یافتن بهترین ترکیب متغیرهای ورودی به مدل استفاده کردند.

رمضان و همکاران (2008) با استفاده از آزمون گاما مهمترین عوامل موثر بر فرآیند تابش روزانه خورشیدی را در حوزه‌ی آبخیز برو¹ در انگلستان تعیین نمودند. مقدم‌نیا و همکاران (2009a) از آزمون گاما برای یافتن بهترین ترکیب متغیرهای ورودی مدل تبخیر روزانه دریاچه‌ی چاه نیمه واقع در استان سیستان و بلوچستان استفاده کردند. مطالعات دیگری در موضوعات مختلف با استفاده از آزمون گاما صورت گرفته است که مقدم‌نیا و همکاران (2009b)،

خشکسالی‌ها همواره امنیت زندگی و اموال مردم ساکن در پایین‌دست هر رودخانه‌ای را تهدید و خسارات فراوانی را به بار آورده است. بنابراین، با مدل‌سازی و پیش‌بینی دقیق دبی جریان رودخانه در فصول مختلف، می‌توان اولین گام را در مدیریت علمی منابع آب‌های سطحی منطقه برداشت. افزون بر این، با ایجاد مخازن ذخیره‌ی آب با ابعاد مناسب روی جریان آب رودخانه علاوه بر کاهش خسارات سیلاب، می‌توان آب مازاد را برای مصارف مختلف در دوره‌های خشکسالی آتی ذخیره نمود. در فرآیند تشکیل رواناب حوزه متغیرهای متعددی دخالت دارد. اثر متقابل این متغیرها باعث می‌شود که این فرآیند به یک پدیده غیرخطی و پیچیده تبدیل شود. از طرفی برای آنکه بتوان یک مدل مناسب برای پیش‌بینی میزان دبی جریان آب رودخانه تهیه نمود، لازم است تا حد امکان تمامی عوامل موثر اندازه‌گیری و در مدل دخالت داده شود. در بیشتر مواقع اندازه‌گیری غلط، ثبت اشتباه داده‌ها، کمبود وقت و یا منابع مالی جهت احداث ایستگاه‌های هیدرومتری و نهایتاً کافی نبودن طول سری داده‌ها باعث می‌شود که محقق قادر به تهیه مدل با دقت قابل قبول نباشد.

تاکنون مطالعات زیادی برای مدل‌سازی و پیش‌بینی دبی جریان رودخانه با روش‌های مختلف انجام شده است. با این حال، هنوز مسائل حل نشده فراوانی در این زمینه وجود دارد. انتخاب بهترین ترکیب متغیرهای ورودی مدل برای پیش‌بینی دقیق دبی جریان رودخانه یکی از آنها می‌باشد. اکثر محققین برای مدل‌سازی از داده‌های دبی رواناب و بارش (با تاخیرهای مختلف زمانی) استفاده کرده‌اند (تیفور و گول‌دال 2006، رمضان و همکاران 2009) و گروه دیگر، علاوه بر متغیرهای مذکور دیگر داده‌های هواشناسی مثل درجه حرارت هوا را برای مدل‌سازی به کار برده‌اند (دورانت 2002). اما این سوال مطرح است که اولاً کدام متغیرها (و با چند تاخیر) برای مدل‌سازی جریان آب رودخانه مناسب می‌باشد و ثانیاً چند درصد از داده‌های

¹ Brue

هیدرولوژیست‌ها می‌باشد. با این حال، این روش در مطالعات مرتبط با جریان آبراه‌های و به ویژه در کشور ما به ندرت استفاده شده است. بنابراین، هدف از انجام این مطالعه یافتن بهینه ترکیب متغیرهای ورودی (شامل بارش و رواناب با تاخیرهای مختلف) در مدل‌سازی، تشخیص تعداد بهینه‌ی داده‌های مربوط به آموزش مدل و پیش‌بینی جریان آب رودخانه در حوزه آبخیز امامه می‌باشد.

مواد و روش‌ها

منطقه‌ی مورد مطالعه

حوزه آبخیز امامه یک حوزه معرف و از زیر حوزه‌های سد لتیان است که به منظور مدل‌سازی جریان آب در این مطالعه انتخاب شده است (شکل 1). این حوزه در طول جغرافیایی $51^{\circ}38' - 51^{\circ}32'$ شرقی و عرض جغرافیایی $35^{\circ}57' - 35^{\circ}51'$ شمالی واقع شده است. حوزه مذکور از شمال به ارتفاعات جنوبی دره لار، از غرب به ارتفاعات اوشان کوه و ارتفاعات شرقی رودخانه جاجرود، از شرق به ارتفاعات راحت‌آباد و کوسا و از جنوب به رودخانه جاجرود و دهکده کمرخانی محدود شده است. در این مطالعه داده‌های روزانه دبی جریان آب رودخانه و داده‌های بارندگی روزانه مربوط به ایستگاه کمرخانی واقع در خروجی حوزه، در طول دوره آماری 1388-1379 مورد استفاده قرار گرفتند. این دوره به دو دلیل نزدیکی این دوره آماری به زمان حال و همخوانی داده‌های بارش و رواناب انتخاب شد. برخی از مشخصات فیزیوگرافی حوزه در جدول 1 نشان شده است.

در این مطالعه هشت متغیر ورودی که برای مدل‌سازی جریان توصیه شده (تیفور و گول‌دال، 2006)، مورد استفاده قرار گرفت. این متغیرها شامل تاخیر یک روزه، دو روزه، سه روزه و چهار روزه در سری مربوط به جریان روزانه و تاخیر یک روزه، دو روزه،

احمدی و همکاران (2009) و پیری و همکاران (2009) از این جمله می‌باشند. در ایران نیز مطالعاتی با استفاده از این روش صورت گرفته است که میتوان به قبایی و همکاران (1388)، صادقی و همکاران (1383) و صادقی و همکاران (1385) اشاره کرد.

رمضان و همکاران (2009) با استفاده از داده‌های روزانه بارش و رواناب و روش هیبرید موجک (NW¹) مقدار رواناب رودخانه را در حوزه آبخیز برو در انگلستان مدل‌سازی کردند. آنها از داده‌های روزانه بارش و رواناب با زمان تاخیرهای مختلف حالت‌های مختلفی را به عنوان متغیرهای ورودی مورد بررسی قرار دادند و با استفاده از آزمون گاما نتیجه گرفتند که سه تاخیر در رواناب و یک تاخیر در بارش بهترین ترکیب ورودی برای مدل‌سازی رواناب رودخانه است. ایشان ترکیب ورودی مذکور را برای تمام روش‌های مدل‌سازی استفاده کردند. آنها همچنین با استفاده از آزمون M به این نتیجه رسیدند که تنها 1056 داده برای استفاده در بخش آموزش مدل جریان آب رودخانه مورد مطالعه در مدل‌سازی کافی است. سپس نتیجه بدست آمده از روش مدل‌سازی NW را با نتایج بدست آمده از روش‌های دیگر مقایسه کردند. نتایج آنها نشان داد که مدل NW از سایر مدل‌ها از دقت بالاتری برخوردار است.

وان جعفر و همکاران (2011)، برای انتخاب متغیر ورودی به منظور تخمین میانه حداکثر جریان سالانه در جنوب غربی انگلستان دو روش آزمون گاما و LOOCV² را استفاده کردند. ایشان نتیجه گرفتند که آزمون گاما توانایی محاسبه واریانس خروجی را با استفاده از ورودی‌ها بدون نیاز به ایجاد ساختار یک مدل خاص را دارد.

به نظر می‌رسد، قابلیت روش آزمون گاما در تعیین ترکیب بهینه متغیرهای ورودی مورد توجه

¹ Neuro Wavlet (NW)

² Leave-On-Out Cross Validation

سه روزه و بدون تاخیر زمانی در سری بارش روزانه
روش آزمون گاما

آزمون گاما یک روش غیر متغیری است و نتایج آن صرف نظر از تکنیک‌های خاص برای مدل‌سازی به کار برده می‌شود. با این تست مقدار میانگین مربعات خطای مدل پیش از استفاده محاسبه و ترکیب بهینه متغیرهای ورودی شناسایی می‌شود. بنابراین، آزمون گاما را می‌توان به عنوان ابزاری مناسب برای یافتن بهترین ترکیب از متغیرهای ورودی به مدل غیر خطی در نظر گرفت. در واقع آزمون گاما تخمینی از حداقل میانگین مربعات خطای آن بخش از داده‌ها است که با

بودند.

مدل خطی نمی‌توان آنها را تعیین کرد. فرض شد

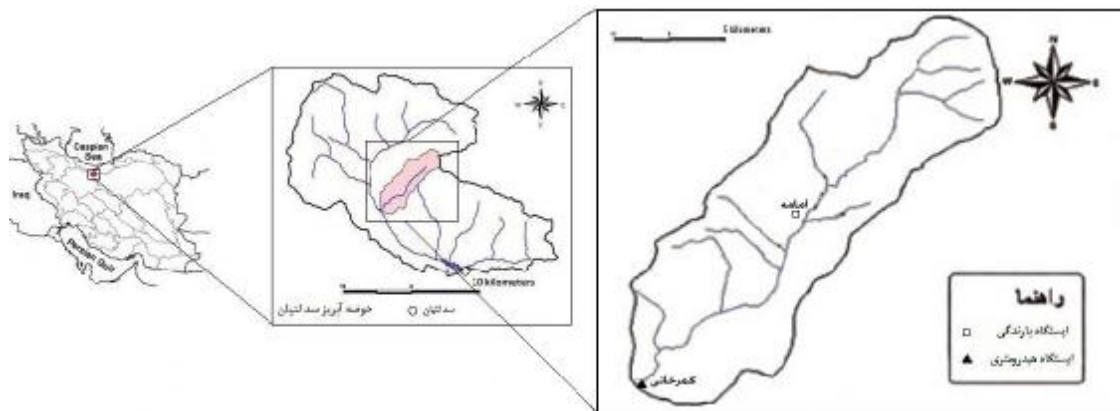
مجموعه داده‌های مشاهداتی به صورت زیر است:

$$\{(x_i, y_i), 1 \leq i \leq M\} \quad [1]$$

که در آن x_i بردار مشاهداتی ورودی (مانند رواناب، بارش، درجه حرارت هوا با تاخیرهای مختلف و هر متغیر دیگر موثر)، y_i خروجی مشاهداتی (مانند میانگین دبی رواناب روزانه) و M تعداد کل مشاهدات می‌باشند.

جدول 1- برخی از مشخصات فیزیوگرافی حوزه آبخیز امامه در محل ایستگاه کمرخانی.

مساحت حوزه (Km ²)	محیط حوزه (Km)	طول آبراهه اصلی (Km)	ضریب شکل	شیب متوسط رودخانه (%)	ارتفاع متوسط (m)
37/2	31	13	1/42	9/2	2650



شکل 1- موقعیت جغرافیایی حوزه آبخیز امامه در استان تهران و کشور.

$$y = f(\mathbf{x}) + r \quad [2]$$

که در آن f یک تابع هموار¹ و r متغیر تصادفی نشان دهنده‌ی خطا می‌باشد. فرض شد میانگین توزیعی

از بردار x_i (در فضای m بعدی) برای مدل‌سازی y_i (در فضای R) استفاده می‌شود. بنابراین، با توجه به فرض فوق رابطه‌ی بین ورودی و خروجی به شرح زیر در نظر گرفته شد:

¹ Smooth Function

متغیر دیگری که معمولاً در آزمون گاما به کار برده می‌شود متغیر بدون بعد V_{ratio} نام دارد که مقدار آن بین صفر و یک بوده و مقدار آن از رابطه 6 محاسبه شد (ایوانس و جونز 2002):

$$V_{ratio} = \frac{\Gamma}{s^2(y)} \quad [6]$$

که در آن $\sigma^2(y)$ واریانس داده‌های دبی جریان خروجی مشاهداتی می‌باشد. هر چه مقدار V_{ratio} به صفر نزدیکتر باشد ترکیب انتخابی برای مدل‌سازی، ترکیب مناسب‌تری است. ثابت شده است که اگر این مقدار از یک کم شود مقدار ضریب تبیین مدل بدست می‌آید (جونز 2004).

اگر فرض شود که N تعداد متغیرهای ورودی موثر بر خروجی مدل باشد در این صورت، تعداد $2^N - 1$ حالت (ترکیب‌های مختلفی از متغیرهای ورودی) وجود دارد که می‌توان از هر کدام برای تهیه مدل استفاده کرد. استفاده از همه ترکیبات ممکن برای یافتن بهترین ترکیب ورودی به مدل بسیار وقت‌گیر و خسته کننده است. بنابراین، در شرایطی که تعداد متغیرهای ورودی زیاد باشند با استفاده از آزمون گاما می‌توان بهترین ترکیب از میان تمام حالات ممکن را تشخیص داد. هم چنین با استفاده از آزمون M تعداد بهینه داده‌های ورودی در بخش آموزش مدل‌سازی تشخیص داده می‌شود. آزمون M در واقع رسم نمودار تغییرات مقدار آماره Γ و یا خطای استاندارد در مقابل تعداد داده‌ها می‌باشد. در نقطه‌ای که این دو نمودار تقریباً به حالت افقی و پایدار می‌رسند آن نقطه تعیین کننده تعداد نقاط کافی جهت استفاده در بخش آموزش مدل‌سازی می‌باشد (جونز 2004).

از آنجایی که داده‌های بارش و رواناب ابعاد متفاوتی دارند و مقادیر آنها اختلاف زیادی با هم داشتند مبادرت به نرمال سازی داده‌ها شد. در این مطالعه داده‌ها بشرح رابطه 7 نرمال شدند.

$$x_N = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \quad [7]$$

که r از آن پیروی می‌کند صفر و واریانس خطا، $var(f)$ کران‌دار است. به این ترتیب، مدل نشان داده شده با رابطه 2 به مدلی هموار با مشتقات جزئی مرتبه اول محدود گردید. آماره گاما (Γ) واریانس بخشی از داده‌ها را که تابع f قادر به توجیه آن نیست نشان می‌دهد.

فرض شد برای بردار ورودی x_i ، مجموعه $N[i,k]$ $(1 \leq k \leq p)$ به نام مجموعه نزدیکترین همسایگی‌ها است. پایه و اساس آزمون گاما بر مبنای مجموعه $N[i,k]$ می‌باشد. $x_{N[i,k]}$ به عنوان k امین نزدیک‌ترین همسایگی‌ها برای بردار مشاهداتی ورودی x_i تعریف شد به طوری که $(1 \leq k \leq p)$ و $(1 \leq i \leq M)$ می‌باشند. p حداکثر تعداد همسایگی‌ها است که معمولاً بین 10 تا 50 در نظر گرفته می‌شود (دورانت 2001). به منظور بدست آوردن مقدار آماره گاما، (Γ)، ابتدا مقدار دلتا، $(\delta_M(k))$ ، با توجه به داده‌های ورودی به شرح زیر محاسبه شد:

$$d_M(k) = \frac{1}{M} \sum_{i=1}^M |x_{N[i,k]} - x_i|^2, \quad 1 \leq k \leq P \quad [3]$$

که در آن نماد $|...|$ به عنوان فاصله اقلیدسی بردار $x_{N[i,k]}$ از بردار همسایگی‌های آن می‌باشد. همچنین مقدار آماره گامای کوچک، $(\gamma_M(k))$ با استفاده از داده‌های خروجی به شرح زیر بدست آمد:

$$g_M(k) = \frac{1}{2M} \sum_{i=1}^M (y_{N[i,k]} - y_i)^2, \quad 1 \leq k \leq P \quad [4]$$

که در آن $y_{N[i,k]}$ مقدار خروجی نظیر k امین نزدیکترین همسایگی برای بردار x_i می‌باشد. به این ترتیب p مقدار برای $\delta_M(k)$ و p مقدار برای $\gamma_M(k)$ بدست آمد. سپس بین p زوج مرتب $\{\delta_M(k), \gamma_M(k)\}$ رابطه رگرسیونی خطی ساده به شرح رابطه 5 برقرار شد. طوری که در آن γ و δ به ترتیب به عنوان متغیرهای وابسته و مستقل در نظر گرفته شدند.

$$g = Ad + \Gamma \quad [5]$$

مقدار آماره Γ در واقع عرض از مبدا خط رگرسیون فوق می‌باشد. A شیب خط رگرسیون بوده که نشان دهنده پیچیدگی مدلی است که از روی داده‌های مشاهداتی ایجاد می‌شود (ایوانس و جونز 2002).

$$\begin{pmatrix} x_{11} & x_{12} & x_{13} & \mathbf{L} & x_{1d} \\ x_{21} & x_{22} & x_{23} & \mathbf{L} & x_{2d} \\ \mathbf{M} & \mathbf{M} & \mathbf{M} & \mathbf{O} & \mathbf{M} \\ x_{p_{\max}1} & x_{p_{\max}2} & x_{p_{\max}3} & \mathbf{L} & x_{p_{\max}d} \end{pmatrix} \begin{pmatrix} m_1 \\ m_2 \\ m_3 \\ \mathbf{M} \\ m_d \end{pmatrix} = \begin{pmatrix} y_1 \\ y_2 \\ \mathbf{M} \\ y_{p_{\max}} \end{pmatrix} \quad [9]$$

اگر ماتریس \mathbf{X} مربعی و غیر منفرد² باشد برای یافتن بردار \mathbf{m} ، از حل معادله زیر بدست می‌آید:

$$\mathbf{m} = \mathbf{X}^{-1}\mathbf{y} \quad [10]$$

در غیر این صورت لازم است تابع هدف زیر کمینه گردد:

$$|\mathbf{Xm} - \mathbf{y}|^2 \quad [11]$$

در این مطالعه نرم افزار win-Gamma برای یافتن بهینه ترکیب متغیرها و تعداد نقاط کافی برای بخش آموزش مدل به کار گرفته شد. همچنین برای ارزیابی عملکرد مدل‌ها از سه آماره ضریب تبیین (R^2)، RMSE و نش- ساتکلیف (NS^3) استفاده شد.

نتایج و بحث

جدول 2 مشخصات آماری داده‌های مورد استفاده در این مطالعه را نشان می‌دهد. همچنین شکل-های 2 و 3 میانگین روزانه بارندگی و رواناب را در طول دوره آماری نشان می‌دهد.

برای مشخص کردن ترکیب بهینه متغیرهای ورودی در مدل ابتدا با استفاده از آزمون گاما مقدار گاما برای حالت استفاده از تمام متغیرهای ورودی بدست آمد. سپس یکی از متغیرهای ورودی حذف و مقدار گاما برای ترکیب مفروض محاسبه شد. آنگاه متغیر مذکور دوباره به مدل وارد و متغیر دیگری حذف شد و مقدار گاما برای ترکیب جدید بدست می‌آمد. این کار برای همه متغیرها تکرار و نتیجه در جدولی ثبت شد. جدول 3 نتایج آزمون گاما را در حوزه امامه نشان می‌دهد. در این جدول ترکیب شماره 1، ترکیبی است که

که در آن x_N مقدار متغیر پس از نرمال‌سازی، x مقدار متغیر مشاهداتی مورد نظر، x_{\max} حداکثر مقدار متغیر مشاهده شده در طی دوره آماری و x_{\min} حداقل مقدار متغیر مشاهده شده در طی دوره آماری می‌باشد. با استفاده از این رابطه داده‌ها بین صفر و یک قرار می‌گیرند.

روش‌های مدل‌سازی

در این مطالعه از دو روش شبکه عصبی مصنوعی (مک‌کلوج 1943) و رگرسیون خطی محلی برای مدل‌سازی استفاده شد. در روش شبکه عصبی مصنوعی از الگوریتم آموزش BFGS¹ (فلچر 1987) برای آموزش داده‌ها استفاده شد.

رگرسیون خطی محلی یکی از روش‌های ساده مدل‌سازی غیر خطی و غیر پارامتریک می‌باشد. از مزایای مهم این روش، مدل‌سازی مطمئن بر اساس تعداد کم داده‌های ورودی و همچنین پیش‌بینی دقیق در محدوده فضای ورودی می‌باشد. این روش از سه نقطه برای مدل‌سازی اولیه و از بقیه نقاط برای بهبود مدل اولیه و ایجاد مدل‌های بعدی استفاده می‌کند. مهم‌ترین بخش روش LLR انتخاب p_{\max} یا آماره تاثیر، می‌باشد. در p_{\max} در واقع تعداد نزدیک‌ترین همسایگی در نظر گرفته شده برای مدل‌سازی خطی می‌باشد. برای یافتن مقدار p_{\max} باید دستگاه معادلات ماتریسی زیر حل شود.

$$\mathbf{Xm} = \mathbf{y} \quad [8]$$

که در آن \mathbf{X} یک ماتریس به ابعاد $p_{\max} \times d$ است، که در آن d تعداد ابعاد بردار ورودی می‌باشد. p_{\max} تعداد نقاط همسایگی برای بردار \mathbf{X}_i ($1 \leq i \leq p_{\max}$)، \mathbf{y} خروجی مشاهداتی که یک بردار ستونی به طول p_{\max} است و \mathbf{m} یک بردار ستونی بوده و مجهول مسئله است. درایه‌های این بردار با استفاده از مشاهدات \mathbf{X} و \mathbf{y} تعیین می‌شود. در نهایت می‌توان رابطه‌ی فوق را به شرح زیر نوشت (مقدم‌نیا و همکاران 2009b).

²Non-singular

³Nash Sutcliffe (NS)

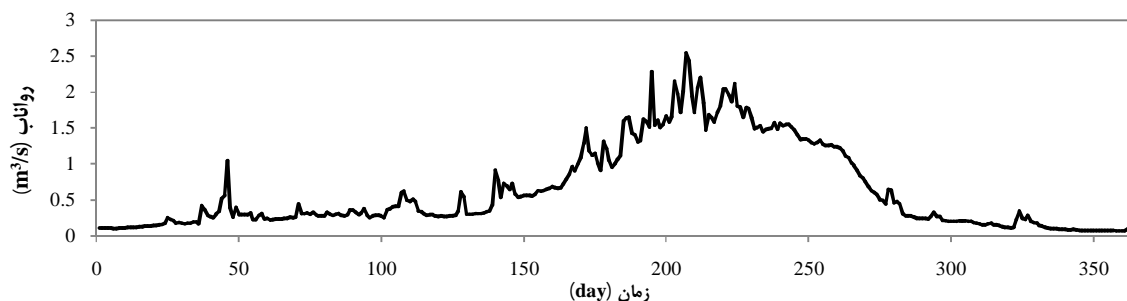
¹Broyden-Fletcher-Goldfarb-Shanno (BFGS)

نمودار مقدار آماره گاما تابعی از تعداد نزدیکترین همسایگی رسم شد. کمترین مقدار گاما، در نمودار مذکور به عنوان حداکثر مقدار نزدیکترین همسایگی انتخاب شد. در این مطالعه تعداد نزدیکترین همسایگی برابر 11 در نظر گرفته شد.

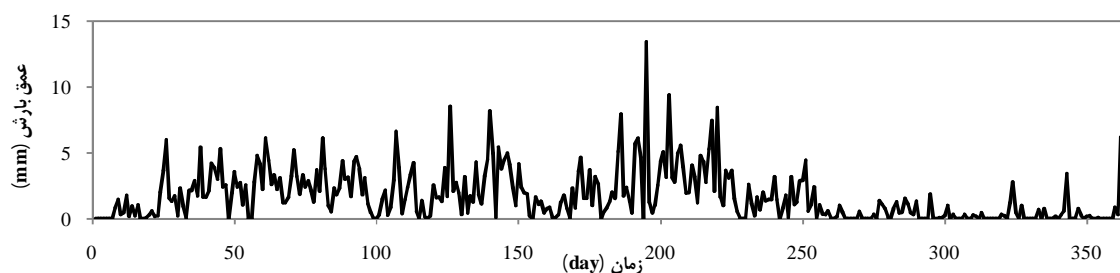
در آن همه متغیرهای ورودی حضور دارد. در ترکیبات دیگر همه متغیرها به جز یکی در مدل حاضر هستند. مثلا در ترکیب شماره 2 همه متغیرها به جز متغیر $R(t)$ (4 حاضر می‌باشند. برای بدست آوردن مقدار بهینه حداکثر مقدار نزدیکترین همسایگی، ابتدا آماره گاما برای مقادیر مختلف نزدیکترین همسایگی محاسبه شد و

جدول 2- مشخصات آماری داده‌های مورد استفاده در حوزه امامه در دوره آماری 1379-1388.

متغیر	واحد	میانگین	انحراف معیار	ضریب تغییرات	حداکثر	حداقل
رواناب	m^3/s	0/635	0/867	1/364	10/8	0/01
بارندگی	mm	1/718	5/36	3/121	79	0/0



شکل 2- میانگین روزانه جریان رودخانه حوزه امامه در ایستگاه هیدرومتری کمرخانی در دوره آماری 1379-1388 (در این شکل روز 1 نشان دهنده اول مهر (اول سال آبی) می‌باشد).



شکل 3- میانگین روزانه بارش حوزه امامه در دوره آماری 1379-1388.

شماره 1 بود، در نظر گرفته شدند. متغیرهای غایب در ترکیب‌های مذکور به عنوان متغیر مهم در تخمین رواناب حوزه امامه تشخیص داده شدند و از آنها برای مدل‌سازی رواناب استفاده گردید. نوری و همکاران (2010) نیز این روش را برای مدل‌سازی به کار برده‌اند.

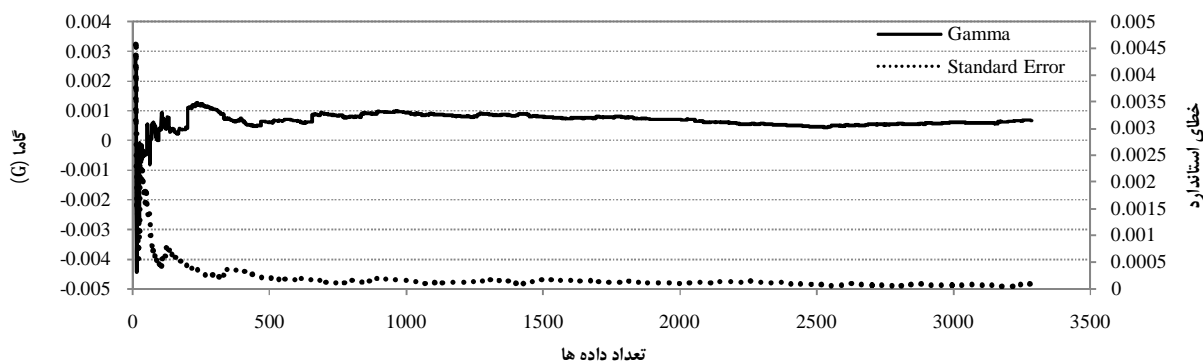
بعد از اینکه مقدار گاما برای تمامی ترکیب‌های جدول 1 بدست آمد مقدار آماره Γ هر یک از ترکیبات شماره 2 تا 9 با مقدار آماره گامای مربوط به ترکیب 1 مقایسه شد. ترکیباتی که در آنها مقدار گامای مدل با ترکیبات مفروض بیشتر از مقدار گامای نظیر ترکیب

ترکیب منتخب، نشان می‌دهد. همان طور که در این شکل مشاهده می‌شود، نمودار آماره گاما (خط پرننگ) و خطای استاندارد (نقطه‌چین) هر کدام از نقطه‌ای به بعد تقریباً هموار می‌شوند. با استفاده از این دو نمودار تعداد 1405 داده برای بخش آموزش مدل در نظر گرفته شد. این رقم کمتر از 40 درصد کل داده‌ها می‌باشد و استفاده از این تعداد پیچیدگی مدل را به طور قابل ملاحظه‌ای نسبت به زمانی که از 70 درصد داده‌ها برای آموزش مدل استفاده می‌شود (مرسوم در شبکه‌های عصبی)، کاهش می‌دهد. رمضان و همکاران (2009)، تعداد 1056 نقطه از 2236 نقطه را (47/3 درصد) با استفاده از آزمون M جهت مدل‌سازی مناسب تشخیص دادند.

به طوری که از جدول 3 می‌توان استنباط کرد، شش ترکیب وجود دارد که مقدار گامای آنها بیش از گامای نظیر ترکیب شماره 1 می‌باشد. بنابراین، متغیرهای غایب این ترکیبات شامل شش متغیر $P(t)$ ، $P(t-1)$ ، $P(t-2)$ ، $P(t-3)$ ، $R(t-1)$ و $R(t-2)$ به عنوان متغیرهای موثر در مدل‌سازی رواناب تشخیص داده شدند. همچنین، متغیر $P(t)$ به عنوان مهم‌ترین متغیر در این ترکیب می‌باشد. زیرا هنگامی که این متغیر از ترکیب ورودی‌های مدل حذف می‌شد مقدار Γ نسبت به ترکیب شماره 1 (یعنی زمانی که همه متغیرها در ترکیب وجود دارند) بیشترین مقدار را به خود اختصاص می‌داد. برای تشخیص اینکه چه تعداد داده برای مدل‌سازی در بخش آموزش مدل کافی است از آزمون M استفاده شد. شکل 4 نتایج آزمون M را با استفاده از متغیرهای استفاده شده در

جدول 3- نتایج آزمون گاما در مدل‌سازی رواناب رودخانه حوزه امامه.

شماره ترکیب	متغیر غایب در مدل	گاما (Γ)	شیب (A)	خطای استاندارد	نسبت V
1	-	0/0006207	0/05616	0/000069	0/09613
2	R(t-4)	0/0005359	0/08001	0/000082	0/083
3	R(t-3)	0/0006054	0/06998	0/000084	0/09377
4	R(t-2)	0/0006427	0/06338	0/000061	0/09954
5	R(t-1)	0/0009275	0/05626	0/000089	0/1436
6	P(t-3)	0/0006825	0/05288	0/000103	0/10571
7	P(t-2)	0/0006837	0/05085	0/000101	0/10589
8	P(t-1)	0/0007210	0/04690	0/000071	0/1116
9	P(t)	0/0009495	0/03109	0/00011	0/1462



شکل 4- نمودار تغییرات مقدار آماره‌ی گاما و خطای استاندارد در مقابل تعداد داده‌ها (آزمون M).

جدول 4- نتایج مدل‌سازی با استفاده از روش رگرسیون خطی محلی با تعداد داده‌های متفاوت در بخش آموزش.

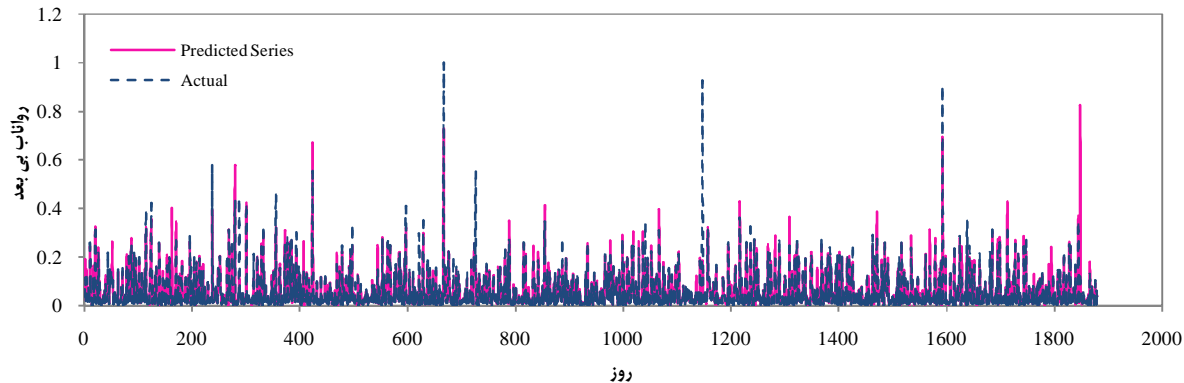
تست		آموزش		تعداد داده‌های مرحله‌ی آموزش
RMSE	R ²	RMSE	R ²	
0/547	0/738	0/119	0/385	500
0/524	0/715	0/153	0/972	750
0/521	0/706	0/15	0/971	1000
0/53	0/72	0/188	0/953	1250
0/385	0/825	0/18	0/954	1500
0/412	0/786	0/18	0/957	1750
0/428	0/762	0/167	0/963	2000
0/489	0/713	0/173	0/961	2250
0/484	0/704	0/167	0/963	2500

که 1500 نقطه برای آموزش مورد استفاده قرار گرفته است. همچنین بعد از 1500 نقطه با افزایش تعداد نقاط در مرحله آموزش کاهش در مقدار R² و افزایش مقدار RMSE دیده می‌شود. بنابراین، تعداد نقاط کافی برای مرحله آموزش در مدل‌سازی باید عددی نزدیک به 1500 باشد. با استفاده از آزمون M تعداد 1405 نقطه برای مرحله آموزش مدل‌سازی انتخاب شد که نتیجه آن با نتیجه مدل‌سازی با تعداد داده‌های مختلف در مرحله آموزش همخوانی دارد.

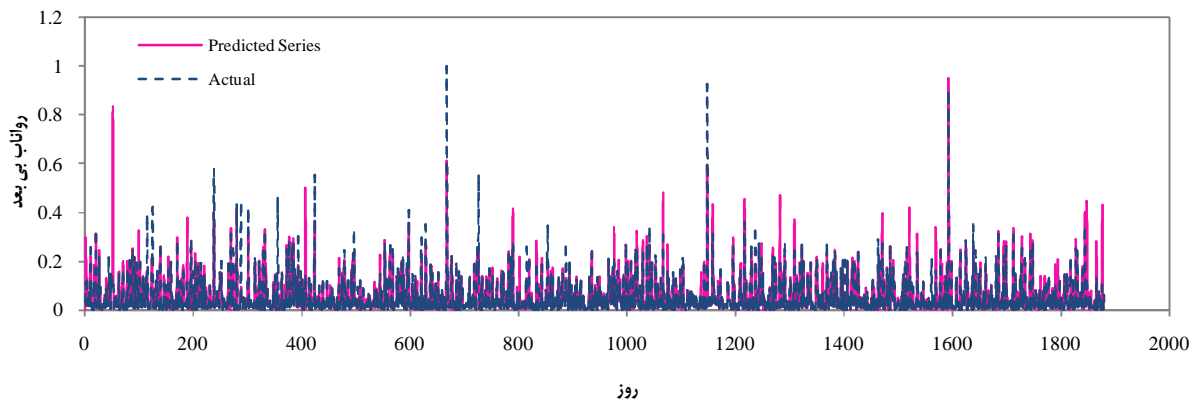
برای اطمینان از نتایج حاصل از آزمون M با استفاده از روش رگرسیون خطی محلی (به عنوان مدل شاهد)، مدل‌سازی با تعداد داده‌های متفاوت در بخش آموزش انجام گرفت. برای این مرحله 500 نقطه از انتهای دوره آماری برای مرحله‌ی ارزیابی در نظر گرفته شد. جدول 4 نتایج مدل‌سازی با تعداد داده‌های متفاوت در مرحله مدل‌سازی را با استفاده از روش LLR نشان می‌دهد. همان‌طور که مشاهده می‌شود بیشترین R² و کمترین RMSE در مرحله تست زمانی حاصل می‌شود

تخمین مقادیر کم و متوسط از دقت خوبی برخوردار است ولی در تخمین مقادیر دبی اوج دقت خوبی ندارد. مدل LLR ایجاد شده در بخش آموزش با توجه به هر سه معیار NS ، R^2 و $RMSE$ از دقت بالاتری نسبت به روش ANN برخوردار است. در صورتی که در بخش تست دقت مدل ایجاد شده با استفاده از روش ANN بالاتر از دقت مدل ایجاد شده با استفاده از روش LLR می‌باشد.

پس از انتخاب بهترین ترکیب از ورودی‌ها و بهینه تعداد داده برای بخش آموزش در مدل‌سازی، رواناب رودخانه حوزه مورد مطالعه با استفاده از دو روش ANN و LLR مدل‌سازی شد. جدول 5 نتایج مدل‌سازی را نشان می‌دهد. شکل‌های 5 و 6 مقادیر سری زمانی بخش تست مدل را به ترتیب برای دو روش ANN و LLR نشان می‌دهد. در این شکل‌ها عدد یک نشان دهنده بیشترین رواناب نرمال شده می‌باشد. همان طور که در شکل 5 مشاهده می‌شود، گرچه مدل بسط داده شده در



شکل 5- سری‌های زمانی مشاهداتی و محاسباتی در بخش تست روش ANN با استفاده از شش متغیر منتخب با روش آزمون گاما و تعداد 1405 داده برای آموزش.



شکل 6- سری‌های زمانی مشاهداتی و محاسباتی در بخش تست روش LLR با استفاده از شش متغیر منتخب با روش آزمون گاما و تعداد 1405 داده برای آموزش.

جدول 5- نتایج مدل‌سازی رواناب حوزه آبخیز امامه با استفاده از شش متغیر منتخب با روش آزمون گاما و تعداد داده برای آموزش با دو روش ANN و LLR.

تست			آموزش			مدل
R ²	RMSE	NS	R ²	RMSE	NS	
0/80	0/63	0/795	0/90	2/56	0/899	ANN(BFGS)
0/68	4/8	0/642	0/96	1/7	0/955	LLR

که صورت گرفته استفاده از آزمون M سبب کاهش تعداد داده‌های مورد استفاده در بخش آموزش شده ولی باز هم به طور قطع نمی‌توان گفت که استفاده از این روش سبب کاهش تعداد داده‌های بخش آموزش می‌شود. از طرفی استفاده از داده زیاد برای آموزش موجب پیچیدگی مدل می‌شود. واقعیت این است که تعداد داده‌های کافی در بخش آموزش به داده‌های ورودی بستگی دارد و ممکن است نتایج در حوضه‌ای دیگر متفاوت با نتایج بدست آمده از این مطالعه برای حوضه آبریز امامه باشد. ولی می‌توان گفت این روش روشی مناسب برای تعیین تعداد نقاط کافی در بخش آموزش مدل‌سازی می‌باشد. واقعیت این است که تاکنون مطالعات کمی در خصوص تعیین نقاط کافی برای بخش آموزش مدل انجام شده است. عدم وجود داده‌های هیدرولوژیکی مشاهداتی به اندازه مورد نیاز در مدل‌سازی یکی از مهمترین مشکلات هیدرولوژیست‌ها می‌باشد. هر چند امروزه با پیشرفت‌های چشمگیر این مشکل (با جمع آوری داده‌های بیشتر) کم رنگ شده است و دانشمندان علم هیدرولوژی باید مطالعاتی را برای تعیین بهترین تعداد نقاط برای مدل‌سازی انجام دهند و روش‌هایی را پیشنهاد کنند. در این مطالعه از M تست برای تعیین بهترین تعداد نقاط برای مدل‌سازی استفاده شد.

نتیجه‌گیری کلی

در این مطالعه بهترین ترکیب متغیرهای ورودی برای مدل‌سازی رواناب روزانه حوزه امامه انتخاب و

مدل رگرسیون خطی محلی ایجاد شده در بخش آموزش عملکرد مطلوبی داشت و R² و RMSE آن به ترتیب برابر 0/96 و 1/7 بدست آمد. در صورتی که در قسمت تست این دو عدد به ترتیب برابر 0/68 و 4/8 بدست آمد و تفاوت زیادی با قسمت آموزش مدل داشت. از خصوصیات مدل‌سازی با روش رگرسیون خطی محلی دقت بالای مدل‌سازی در فضای آموزش می‌باشد (رمضان و همکاران 2009). این نتیجه با یافته‌های مطالعه حاضر در حوزه امامه همخوانی دارد. در مدل ایجاد شده با استفاده از روش ANN با اینکه مقادیر آماره‌های R² و NS در بخش آموزش بیشتر از بخش تست مدل بودند ولی تفاوت آنها نسبت به مدل LLR کمتر بود.

یکی از معایب استفاده از ANN تعیین ترکیب ورودی با استفاده از سعی و خطا می‌باشد که نیاز به صرف وقت زیادی دارد. در این مطالعه با استفاده از روش آزمون گاما قبل از بسط هر گونه مدل خاصی، ترکیب بهینه متغیرهای ورودی انتخاب شد. استفاده از این روش منجر به کاهش مدت زمان لازم برای مدل‌سازی می‌شود. یکی دیگر از معایب ANN تعیین بخشی از داده‌ها جهت استفاده در بخش مدل‌سازی است. گرچه به طور معمول، حدود 70 درصد داده‌ها برای استفاده در این بخش اختصاص می‌یابد ولی آزمون M مشخص کرد که در مورد رودخانه حوزه امامه تنها حدود 40 درصد داده‌ها جهت استفاده در بخش آموزش مدل‌سازی کافی می‌باشد. هر چند تاکنون در مطالعاتی

مدل شد و نتایج آن با استفاده از معیارهای R^2 ، RMSE و NS مورد ارزیابی قرار گرفت. نتایج نشان داد که مدل LLR در قسمت آموزش از دقت بیشتری نسبت به مدل ANN برخوردار است. در صورتی که در قسمت تست، مدل ایجاد شده با استفاده از ANN دقت بیشتری نسبت به مدل LLR داشت.

پیشنهاد می‌شود روش آزمون گاما برای مدل-سازای رواناب حوزه‌های دیگر ایران نیز بکار رود. انتخاب ترکیب بهینه متغیرهای ورودی به مدل با روش آزمون گاما انجام و نتایج با دیگر روش‌های شناخته شده مرسوم مثل رگرسیون خطی چندگانه (با گزینه‌ی گام به گام) و یا روش پروکراستس (دین‌پژوه و همکاران 2004) مقایسه گردد.

سپاسگزاری

نویسندگان مقاله از داوران محترم که با ارائه نکات ارزشمند موجب ارتقای کیفیت مقاله گردیدند، تشکر و قدردانی می‌نمایند.

تعداد نقاط کافی برای قسمت آموزش مدل‌سازی با روش آزمون گاما تشخیص داده شد. برای این کار ابتدا سری‌های مورد استفاده در ورودی مدل، که داده‌های بارش و رواناب روزانه با تاخیرهای مختلف بودند، نرمال شدند. سپس از بین 8 متغیر ورودی ترکیب بهینه متغیرها با استفاده از آزمون گاما بدست آمد. ترکیب منتخب شامل شش متغیر $P(t)$ ، $P(t-1)$ ، $P(t-2)$ ، $P(t-3)$ ، $R(t-1)$ و $R(t-2)$ بود و برای مدل‌سازی با استفاده از دو روش شبکه عصبی مصنوعی و رگرسیون خطی محلی مورد استفاده قرار گرفت. بنظر می‌رسد، با استفاده از آزمون گاما دیگر نیازی به مدل‌سازی با استفاده از سعی و خطا در ترکیب ورودی برای یافتن بهترین مدل نیست. با استفاده از گاما سعی شد تا بهترین ترکیب برای مدل‌سازی انتخاب شود. هدف بعدی یافتن تعداد داده کافی برای بخش آموزش مدل‌سازی بود که در این مطالعه با استفاده از آزمون M تعداد 1405 داده برای آموزش مدل مناسب تشخیص داده شد. بعد از یافتن بهترین ترکیب ورودی و تعداد نقاط کافی برای استفاده در بخش آموزش مدل‌سازی، رواناب روزانه رودخانه حوزه آبخیز امامه با استفاده از دو روش ANN و LLR

منابع مورد استفاده

قبائی سوق م، مساعدی ا، حسام م و هزارجریبی ا، 1389. ارزیابی تاثیر پیش پردازش پارامترهای ورودی به شبکه عصبی مصنوعی با استفاده از روش رگرسیون گام به گام و آزمون گاما به منظور تخمین سریع‌تر تبخیر و تعرق روزانه. نشریه آب و خاک (مشهد). شماره 24، جلد 3. صفحه‌های 610 تا 624.

صادقی ح، نیکپور ع و ایوب زاده ع، 1383. تخمین رسوب روزانه با استفاده از مدل سازی دینامیک در حوزه آبخیز کسلیان. مجله منابع طبیعی ایران. شماره 57. صفحه‌های 391 تا 402.

صادقی ح، نجفی د و وفاخواه م، 1385. تحلیل منطقه‌ای برآورد رسوب معلق در حوزه اصفهان و سیرجان. تحقیقات منابع آب ایران. سال 2، شماره 3. صفحه‌های 51 تا 65.

بی نام. 1379. گزارش آماری حوضه معرف امامه (سال‌های 1349 تا 1379). مرکز تحقیقات منابع آب، وزارت نیرو، تهران.

Ahmadi A, Han D, Karamouz M and Remesan R, 2009. Input data selection for solar radiation estimation. Hydrological Processes 23: 2754-2764.

Corcoran J, Wilson I and Ware J, 2003. Predicting the geo-temporal variations of crime and disorder. International Journal of Forecasting 19: 623-634.

- Dinpashoh Y, Fakheri-Fard A, Moghaddam M, Jahanbakhsh S and Mirnia M, 2004. Selection of variables for the purpose of regionalization of Iran's precipitation climate using multivariate methods. *Journal of Hydrology* 297: 109-123.
- Durrant PJ, 2001. win-GammaTM: A non-linear data analysis and modeling tool with applications to flood prediction. PhD Thesis, Department of Computer Science, Cardiff University, Wales, UK.
- Evans D, 2001. Data derived estimations of noise using near neighbor distance distributions. PhD Thesis, Cardiff University, Wales, U.K.
- Evans D and Jones AJ, 2002. A proof of the gamma test. Pp. 2759-2799. *Proceedings of Royal Society, Series A* 458 (2027).
- Fletcher R, 1987. *Practical Method of Optimization*. 2nd ed. Wiley: New York.
- Jones AJ, 2004. New tools in non-linear modeling and prediction. *Computational Management Science* 1: 109-149.
- Koncar N, 1997. Optimization methodologies for direct inverse neurocontrol. PhD Thesis, Department of Computing, Imperial College of Science, Technology and Medicine, University of London, London.
- McCulloch WS and Pitts W, 1943. A logical calculus of the ideas immanent in nervous activity. *Bull Math Biophys* 5:115-33.
- Moghaddamnia A, Ghafari Gousheh M, Piri J and Han D, 2008. Evaporation estimation using support vector machines technique. *World Academy of Science, Engineering and Technology* 43: Pp. 14-22.
- Moghaddamnia A, Ghafari Gousheh M, Piri J, Amin S and Han D, 2009a. Evaporation estimation using artificial neural networks and adaptive neuro-fuzzy interface system techniques. *Advances in Water Resources* 32: 88-97.
- Moghaddamnia A, Remesan R, Hassanpour Kashani M, Mohammadi M, Han D and Piri J, 2009b. Comparison of LLR, MLP, Elman, NNARX and ANFIS Models - with a case study in solar radiation estimation. *Journal of Atmosphere and Solar-Terrestrial Physics* 71: 975-982.
- Noori R, Karbassi A and Sabahi MS, 2009. Evaluation of PCA and gamma test techniques on ANN operation for weekly solid waste prediction. *Journal of Environmental Management* 91: 767-771.
- Piri A, Amin S, Moghaddamnia A, Keshavarz A, Han D and Remesan R, 2009. *Journal of Hydrologic Engineering* 14(8): 803-811.
- Remesan R, Shamim MA and Han D, 2008. Model data selection using gamma test for daily solar radiation estimation. *Hydrological Processes* 22: 4301-4309.
- Remesan R, Shamim MA, Han D and Mathew J, 2009. Runoff prediction using an integrated hybrid modeling scheme. *Journal of Hydrology* 372: 48-60.
- Stefansson A, Koncar N and Jones AJ, 1997. A note on the gamma test. *Neural Computing and Applications* 5 (3): 131-133.
- Tayfur G and Guldal V, 2006. Artificial neural networks for estimating daily total suspended sediment in natural streams. *Nordic Hydrology* 37 (1): 69-79.
- Wan Jaafar WZ, Liu J and Han, 2011. Input variable selection for median flood regionalization. *Water Resources Research* 47: 1-18.
- Zhang YX, 2007. Artificial neural networks based principal component analysis input selection for clinical pattern recognition analysis. *Talanta* 73: 68-75.
- Zhang YX, Li H, Hou A and Haval J, 2006. Artificial neural networks based on principal component analysis input selection for quantification in overlapped capillary electrophoresis peaks. *Chemo Metrics and Intelligent Laboratory Systems* 82: 165-175.