

پیش‌بینی جریان رودخانه با روش جمعی احتمالاتی نزدیکترین همسایگی

هادی ثانی خانی^{1*} - یعقوب دین پژوه² - محمد علی قربانی³ - مهدی ضرغامی⁴

تاریخ دریافت: 1393/03/17

تاریخ پذیرش: 1393/08/05

چکیده

پیش‌بینی احتمالاتی متغیرهای هیدرولوژیکی در سال‌های اخیر از جمله جریان رودخانه مورد توجه محققین بوده است. پژوهش حاضر دو روش مختلف احتمالاتی را برای پیش‌بینی جریان رودخانه استفاده می‌کند. سپس میزان عدم قطعیت ناشی از پیش‌بینی به صورت کمی معین می‌شود. تخمین - گرهای خطی یکی از روش‌ها و در دیگری الگوریتم نزدیکترین همسایگی استفاده شد. دیروزانه به مدت دوازده سال در دو ایستگاه آسنجیدیز و ماشین به ترتیب واقع در خروجی‌حوضه‌های آبریز باراندوزچای در استان آذربایجان غربی رود زرد در استان خوزستان استفاده شد. شش سال اول داده‌ها جهت برازش، سه سال بعدی برای واسنجی و 3 سال پایانی جهت آزمون الگوها استفاده شدند. ترکیبات مختلفی از داده‌های ثبت شده دبی روزانه به عنوان الگوی ورودی برای پیش‌بینی دبی روزانه استفاده شد. به کارگیری روش‌های مورد استفاده به صورت جمعی (به منظور انتخاب پارامترهای بهینه الگو) منجر به بهبود و افزایش دقت الگو در پیش‌بینی جریان رودخانه شد. از شاخص‌های آماری شامل ضریب همبستگی، ریشه میانگین مربعات خطا و ضریب کارایی ناش - ساتکلیف برای ارزیابی عملکرد الگوها استفاده شدند. نتایج به دست آمده در این تحقیق کارایی و اعتبار روش‌های مورد استفاده را تأیید نمود. همچنین نتایج به دست آمده نشان داد که روش جمعی مبتنی بر الگوریتم نزدیکترین همسایگی عملکرد بهتری در مقایسه با روش جمعی غیرخطی احتمالاتی دارد و از عدم قطعیت کمتری در پیش‌بینی برخوردار است. مقدار ضریب کارایی ناش - ساتکلیف در روش جمعی مبتنی بر نزدیکترین همسایگی در ایستگاه‌های دیزج و ماشین در دوره آزمون الگو به ترتیب برابر 0/91 و 0/93 به دست آمدند.

واژه‌های کلیدی: باراندوزچای، دیزج، رود زرد، ماشین، متغیرهای هیدرولوژیکی

مقدمه

متروپولیس⁵ (11) و روش تخمین عدم قطعیت درست‌نمایی تعمیم یافته⁶ اشاره کرد. گروه دوم، روش‌هایی بر اساس پردازش خطاهای پیش‌بینی موجود در داده‌های گذشته برای تعیین توزیع احتمالاتی حاکم بر مقادیر و داده‌های آتی نظیر مطالعات مونتکاری و براس (13) است. گروه سوم چگونگی انتشار عدم قطعیت از پیش‌بینی بارش به پیش‌بینی دبی رودخانه را ارزیابی می‌کنند. می‌توان به سیستم پیش-بینی بی‌زی⁷ که توسط کریستوفوویز (10) ارائه شده، اشاره کرد. روش پیش‌بینی غیرخطی شباهت‌هایی به روش‌های گروه دوم دارد و در آن از مدل‌سازی فضای حالت استفاده می‌شود. روش پیش‌بینی غیرخطی اولین بار در تحلیل سری‌های زمانی غیرخطی به کار گرفته شد (8). این روش در مطالعات هیدرولوژی و منابع آب نیز توسط محققین مختلف استفاده شده است (9 و 20). نتایج به دست آمده با این روش حاکی از قابلیت بالای آن در مدل‌سازی فرآیند

پیش‌بینی احتمالاتی متغیرهای هیدرولوژیکی مانند جریان رودخانه در سال‌های اخیر همواره مورد توجه محققین بوده است. هدف در رویکرد احتمالاتی پیش‌بینی با یک اطمینان قابل قبول است (4). عدم قطعیت موجود در پیش‌بینی یک مساله اساسی در مطالعات هیدرولوژیکی است. تحلیل عدم قطعیت می‌تواند در ارزیابی دقت، کیفیت پیش‌بینی، انجام عملیات آبی و تصمیم‌گیری‌های متناظر با مقادیر مورد انتظار از پدیده مورد نظر مفید واقع شود (23).

روش‌های احتمالاتی (تعیین عدم قطعیت در پیش‌بینی) در سه گروه کلی طبقه‌بندی می‌شوند. شیوه باز نمونه‌گیری در گروه اول استفاده می‌شود. این روش برای ارزیابی عدم قطعیت پارامترها و مدل در پیش‌بینی رواناب به کار می‌رود. از آن جمله می‌توان به الگوریتم

1. 2 و 3- دانشجوی دکتری و دانشیاران گروه مهندسی منابع آب، دانشکده کشاورزی، دانشگاه تبریز

* - نویسنده مسئول: (Email: hsanikhani12@gmail.com)

4- دانشیار گروه مهندسی آب، دانشکده مهندسی عمران، دانشگاه تبریز

5- Metropolis Algorithm

6- Generalized Likelihood Uncertainty Estimation (GLUE)

7- Bayesian Forecasting System (BFS)

که پیش‌بینی‌های جمعی انجام شده، اطلاعات با ارزش و مفیدی برای مدیریت و برنامه‌ریزی بهینه منابع در حوضه ارائه می‌دهد. هر و کریستوفوویز (7) با روش پیش‌بینی جمعی بی‌زی⁶ به تخمین ارتفاع آب رودخانه پرداختند.

آلفیری و همکاران (2) از شبیه‌سازی جمعی برای پیش‌بینی سریع سیلاب‌های ناگهانی جنوب سوئیس استفاده کردند. آن‌ها آستانه تجاوز⁷ مقادیر دبی پیش‌بینی شده را با توزیع احتمالاتی گاما برآزش دادند و عملکرد روش به کار گرفته شده را در پیش‌بینی سیلاب مناسب عنوان نمودند. فان و همکاران (5) از روش پیش‌بینی جمعی به منظور برآورد مدل‌سازی جریان رودخانه‌ای در کشور برزیل استفاده کردند. نتایج به دست آمده نشان داد که استفاده از روش جمعی اهمیت بسزایی در بهره‌برداری از مخزن و کنترل سیلاب پایین دست دارد.

بررسی مطالعات قبلی نشان داد که دامنه مطالعات انجام شده در خصوص روش‌های جمعی در زمینه منابع آب و هیدرولوژی محدود بوده و در سال‌های اخیر در حال توسعه و گسترش است. لذا هدف از این تحقیق استفاده از روش پیش‌بینی جمعی در تخمین مقدار جریان رودخانه است. روش پیش‌بینی جمعی در قالب دو روش عملگرهای خطی و الگوریتم نزدیکترین همسایگی مدنظر است.

مواد و روش‌ها

روش پیش‌بینی جمعی احتمالاتی غیر خطی

مجموعه‌ای متشکل از n سری زمانی که اعضای آن بشرح $\{y_1(t_i)\}, \{y_2(t_i)\}, \dots, \{y_N(t_i)\}$ نمایش داده می‌شوند و هر کدام از اعضای مجموعه خود دارای N مشاهده t_i ($i = 1, \dots, N$) در فواصل زمانی مساوی Δt هستند در نظر گرفته شد. رابطه چند متغیره با روش تاخیر زمانی مطابق رابطه 1 استفاده شد.

$$y(t) = \left\{ \begin{array}{l} y_1(t), y_1(t - \gamma_1), \dots, y_1(t - (m_1 - 1)\gamma_1), \\ y_2(t), y_2(t - \gamma_2), \dots, y_2(t - (m_2 - 1)\gamma_2), \dots, \\ y_N(t), y_N(t - \gamma_N), \dots, y_N(t - (m_N - 1)\gamma_N) \end{array} \right\} \quad (1)$$

هر ردیف نشان دهنده وضعیت سیستم پویا و زمان t با مولفه‌های m_w برای متغیر w ($w = 1, \dots, N$) است. در این رابطه تاخیرهای زمانی با ضرب مقادیر w ($w = 1, \dots, N$) در Δt به دست آمدند. همچنین بعد فضای برداری P برابر بعد محاط یعنی $m = (m_1 + m_2 + \dots + m_N)$ می‌باشد. اگر هدف پیش‌بینی جریان رودخانه باشد، $y_1(t_i)$ می‌تواند معرف سری زمانی دبی جریان بوده و سایر متغیرها نیز می‌توانند بارش، دما، مشخصات

بارش-رواناب است. نتایج قابل قبولی را نیز در پیش‌بینی جریان رودخانه داشته است (12). ویژگی اصلی روش پیش‌بینی غیرخطی به قابلیت آن در ارائه روند غیرخطی پدیده مورد بررسی و انطباق آن با مقادیر حداکثر و حداقل است.

روش‌های ناپارامتری نظیر رویکرد نزدیکترین همسایگی¹ (K-NN) برای تشخیص الگو و پیش‌بینی همواره مورد توجه محققین بوده است. کمبر و همکاران (9) از روش K-NN جهت پیش‌بینی دبی جریان روزانه رودخانه استفاده نمودند. مقایسه نتایج روش K-NN با مدل خود همبسته میانگین متحرک² (ARMA) حاکی از این بود که عملکرد روش K-NN به مراتب بهتر از ARMA است. شارما و لال (21) یک روش ناپارامتری را برای شبیه‌سازی بارش روزانه برگرفته از الگوریتم K-NN ارائه دادند. روش پیشنهادی برای یک دوره 123 ساله بارش روزانه در سیدنی استرالیا بررسی و عملکرد آن موفقیت آمیز گزارش شد. وئو و چائو (24) عملکرد مدل‌های مختلف شامل مدل شبکه‌های عصبی مصنوعی³ (ANN)، ARMA و K-NN را در پیش‌بینی جریان ماهانه در چندین رودخانه از کشور چین بررسی کردند. نتایج به دست آمده حاکی از آن بود که عملکرد K-NN در مقایسه با سایر روش‌ها بهتر است. عزمی و عراقی نژاد (1) به منظور پیش‌بینی جریان رودخانه در حوضه بالادست سد زاینده‌رود از روش K-NN استفاده کردند. آن‌ها این روش را برای سری‌های تاریخی بلندمدت مناسب دانستند.

روش‌های جمعی⁴ روش‌های نوین پیش‌بینی هستند که در چند سال اخیر در علوم مهندسی آب توسعه یافته‌اند. ارائه عدم قطعیت در پیش‌بینی‌های انجام شده، علت توسعه این روش‌ها است. این روش‌ها بر خلاف پیش‌بینی‌های نقطه‌ای عملکردی محافظه‌کارانه در انجام پیش‌بینی دارند. این مساله در مطالعات هیدرولوژی و منابع آب که پیش‌بینی‌های دبی جریان آب همواره با عدم قطعیت همراه است، حائز اهمیت است. روش پیش‌بینی غیرخطی می‌تواند در قالب پیش‌بینی جمعی مورد استفاده قرار گیرد.

تاما و همکاران (22) روش جمعی احتمالاتی غیرخطی را برای پیش‌بینی جریان رودخانه ارائه نمودند. این روش احتمالاتی برای بهینه کردن پارامترها و افزایش قدرت الگو در مرحله واسنجی ارائه شد. نتایج به دست آمده حاکی از کارایی و قابلیت روش پیشنهادی بود. ری‌گوندا و همکاران (18) چارچوبی برای پیش‌بینی‌های جمعی با چند مدل ارائه و آن را برای پیش‌بینی جریان‌های فصلی در حوضه رودخانه گانیسون⁵ در ایالات متحده بکار گرفتند. آن‌ها عنوان کردند

- 1- K-nearest neighbor (K-NN)
- 2- Autoregressive moving average (ARMA)
- 3- Artificial neural networks (ANN)
- 4- Ensemble methods
- 5- Gunnison

6- Bayesian Ensemble Forecast
7- Thresholds Exceedance

یا $[\theta] \sigma^2$ برابر است. توزیع تجربی مقادیر $(t_0 + T)_k$ (تایه ازای $k = 1, \dots, r$) ارائه‌گر پیش‌بینی احتمالاتی است که محدوده‌های اطمینان شامل چندک‌های $\beta/2$ و $1 - \beta/2$ این توزیع احتمالاتی متناظر با سطح معنی‌داری فرض شده (β) می‌باشد. جزئیات بیشتر این روش در مطالعات تاما و همکاران (22) ارائه شده است.

یک پیش‌بینی جمعی شامل مجموعه‌ای است که اعضای آن مقادیر پیش‌بینی شده که همه آن‌ها در یک زمان مشخص برآورد شده‌اند، می‌باشند. این پیش‌بینی‌ها شامل کلیه حالت‌های ممکن می‌باشند که عدم قطعیت را در پیش‌بینی لحاظ می‌کنند. از مزایای پیش‌بینی‌های جمعی برآورد احتمال رویدادهای مختلف و تخمین بازه اطمینان موجود در پیش‌بینی می‌باشد (19). برابری پیش‌بینی غیرخطی جمع‌یلازم است پارامترهای r ، m_w و γ_w به‌طور صحیح انتخاب شوند. یک راه حل عملی جهت تعیین بهترین مقادیر برای r ، m_w و γ_w روش آزمون و خطا می‌باشد (15). جستجو برای یافتن پارامترهای بهینه با استفاده از روش پیش‌بینی جمعی به شرح زیر انجام شد (17).

در گام اول داده‌ها به سه بخش تقسیم شدند. اول مجموعه داده‌های برآزش که شامل بخشی از داده‌ها در فضای P هستند که نقاط همسایگی به ازای هر نقطه شروع در آن جستجو و شناسایی می‌شوند. بخش دوم داده‌ها شامل داده‌های واسنجی هستند که جهت ارزیابی توانایی مدل در پیش‌بینی (بر مبنای پارامترهای انتخابی) و تعیین بهترین ترکیب پارامترها در الگوبندی می‌باشند. بخش سوم نیز شامل بخشی از مشاهدات هستند که برای آزمون الگوی بسط داده شده، به کار می‌روند و روش پیش‌بینی با این بخش از داده‌ها ارزیابی می‌شود. در گام دوم، محدوده‌ای از مقادیر مختلف برای هر پارامتر انتخاب می‌شود. محدوده تغییرات r از حداقل مقدار آن یعنی $m+1$ تا کل نقاط موجود در مجموعه داده‌های برآزش می‌باشد. مقادیر پارامترهای بازسازی شامل m_w و γ_w نیز در محدوده بین 1 و یک مقدار بزرگ نظیر زمان تمرکز حوضه انتخاب می‌شود.

در گام سوم، روش پیش‌بینی غیرخطی برای شبیه‌سازی داده‌های واسنجی (با مجموعه پارامترهای آزمایشی) به کار گرفته شد و عملکرد مدل پیش‌بینی (با یک شاخص خطا نظیر ریشه میانگین مربعات خطا) ارزیابی شد. در گام آخر نتایج با توجه به خطای پیش‌بینی به صورت صعودی مرتب شده و تعداد N_s مورد از نامزدهای مختلف برای ترکیبات پارامترها (دارای کمترین خطا و بهترین پیش‌بینی در مجموعه داده‌های واسنجی) انتخاب شده و در مجموعه داده‌های آزمون استفاده شدند. همچنین پیش‌بینی نهایی منفرد (میان‌ه پیش‌بینی‌های N_s) مشخص شد. در این مطالعه مقدار میان‌ه (به لحاظ ارائه تخمین بهتر و نتایج قابل قبول‌تر) که نسبت به مقدار میانگین ارجحیت دارد (6) برای پیش‌بینی نهایی منفرد انتخاب شد.

فیزیکی حوضه نظیر رطوبت خاک، نوع پوشش گیاهی یا مقدار دبی جریان در سایر شاخه‌های رودخانه یا ایستگاه‌های آبنجی بالادست باشند.

تغییرات زمانی سیستم می‌تواند به صورت نگاشتی از $y(t)$ به $y(t+T)$ بیان شود. T زمان پیش‌بینی است که بصورت مضربی صحیح از Δt در نظر گرفته‌شود. به منظور تخمین اولین مولفه $y(t+T)$ رابطه 2 بشرح زیر در نظر گرفته شد.

$$y(t+T) = \tilde{y}(t+T) + \theta(t+T) \quad (2)$$

$$= F(y(t)) + \theta(t+T)$$

در این رابطه $y(t+T)$ مقدار مشاهداتی، $\tilde{y}(t+T)$ مقدار تخمینی با استفاده از نگاشت قطعی $R^m \rightarrow R$ و θ معرف خطای پیش‌بینی است. خطای پیش‌بینی شامل خطای مشاهداتی، اغتشاشات نظیر وجود نیروهای خارجی ناشناخته، خطای پارامتریک ناشی از فرآیند بازسازی فضای P و خطای الگوسازی در قالب نگاشت F می‌باشد.

وجود رفتارهای غیرخطی و عدم همگنی تغییرات در نواحی مختلف از فضای P ، مانع اصلی در تعریف کلی نگاشت F (توصیف کننده پویایی فضای P) است. این مساله موجب پیچیدگی صورت نگاشت می‌شود که تنها توسط برخی ابزارهای خاص نظیر شبکه‌های عصبی یا توابع پایه شعاعی قابل الگوسازی است. یک روش جایگزین در این راستا این است که ناحیه P به نواحی همگن پویا تجزیه شود و برای هر زیر مجموعه از نگاشت‌های ناحیه‌ای به صورت جداگانه جهت تقریب استفاده شود. این تقسیم‌بندی موجب می‌شود تا پویایی پیچیده با استفاده از نگاشت ناحیه‌ای در قالب الگوی خطی ساده ارائه شود و سبب افزایش دقت پیش‌بینی و کاهش حساسیت الگو به نوسانات و تغییرات متغیر مورد نظر شود (16).

محدوده نگاشت‌های ناحیه‌ای با نقاط $(k = 1, \dots, r)$ $y(t_k) =$ $(t_k < t_0)$ تشکیل می‌شود. این نقاط نزدیکترین نقاط به نقطه شروع $y(t_0)$ واقع در فضای P هستند. منظور از نقطه شروع نقطه‌ای است که از آن جا پیش‌بینی شروع می‌شود. نقاط $y(t_k)$ معرف نقاط همسایگی $y(t_0)$ می‌باشند. تعداد همسایگی‌ها (r) نقش مهمی در روش پیش‌بینی دارند و تعداد آن‌ها می‌تواند از چندین نقطه تا کلیه نقاط موجود در فضای P را شامل شود (3).

با توجه به رابطه 2، پیش‌بینی قطعی $\tilde{y}(t_0 + T)$ تخمینی از مقدار مشاهداتی است که تحت تاثیر عدم قطعیت $\theta(t_0 + T)$ می‌باشد. نکته اصلی در روش احتمالاتی این واقعیت است که چندین حالت مختلف از $\tilde{y}(t_0 + T)$ نظیر $\tilde{y}(t_0 + T)_k$ ($k = 1, \dots, r$) می‌تواند به دست آمده و به پیش‌بینی‌های قطعی به دست آمده اضافه گردد. به ازای هر یک از حالت‌ها، یک خطای پیش‌بینی متناظر با آن نظیر $\theta_k = y(t_0 + T) - \tilde{y}(t_0 + T)_k$ وجود دارد. با در نظر گرفتن $y(t_0 + T)$ به عنوان یک مقدار مشاهداتی قطعی، واریانس پیش‌بینی یا $[\tilde{y}(t_0 + T)] \sigma^2$ دقیقاً با واریانس خطای پیش‌بینی

رابطه زیر استفاده می‌شود.

$$y_{pred}^{nn} = \frac{\sum_{i \in I_{nn}} y_i}{|I_{nn}|} \quad (5)$$

که در آن I_{nn} مجموعه‌ای اندیس گزار است و اندیس‌داده‌های مشاهداتی است که در داخل دایره‌ای به شعاع b از نقطه پرسش قرار دارند. $|I_{nn}|$ تعداد عضوهای مجموعه I_{nn} مطابق رابطه 6 است.

$$|I_{nn}| = \{i: \|x_i - x_0\| \leq b\} \quad (6)$$

بهبود عملکرد این روش، به انتخاب پارامترهای b (شعاع همسایگی) و l (تعداد تاخیرها در پارامترهای ورودی) بستگی دارد. (14)

روش NNPE شباهت زیادی به روش غیرخطی دارد، اما تفاوت اصلی آن‌ها در این است که در روش غیرخطی از تخمین‌گرهای ناحیه‌ای خطی استفاده شده است، حال آن که در ساختار روش NNPE از روش نزدیکترین همسایگی به عنوان پایه و ساختار اصلی روش استفاده می‌شود. ابتدا سری زمانی داده‌ها در روش NNPE به دو بخش آموزش و آزمون تقسیم می‌شود، سپس داده‌های بخش آموزش نیز به دو بخش جدا از هم واسنجی³ و غیر واسنجی⁴ تقسیم می‌شوند. ابتدا یک نقطه پرسش مشخص در روش نزدیکترین همسایگی از بین اعضای مجموعه داده‌های واسنجی انتخاب و پارامترهای مدل (شعاع همسایگی و تعداد تاخیرها) به دلخواه انتخاب می‌شود. آنگاه مقدار دبی خروجی پیش‌بینی و عملکرد مدل مذکور با توجه به مشاهدات مجموعه غیر واسنجی مورد ارزیابی می‌شود. سپس پارامترهای مدل را تغییر داده و عملکرد مدل ارزیابی می‌شود. الگوهای مورد آزمون با توجه به معیار ارزیابی ($RMSE$) از بزرگ به کوچک مرتب شدند. یک تعداد مشخصی (در این مطالعه صد الگوی اول) از الگوهایی که دارای بهترین عملکرد باشند به عنوان یک گروه الگوی پیش‌بینی کننده برتر برای نقطه پرسش مذکور در نظر گرفته شد. با هر الگوی یک تخمین انجام شد و میانه آن‌ها به عنوان جواب (منفرد) در نظر گرفته شد. با توجه به وجود عدم قطعیت در روش جمعی این روش ماهیت تصادفی دارد. در ادامه با توجه به مقادیر پیش‌بینی شده از صد الگوی برتر صدک $2/5$ و $97/5$ از بین پیش‌بینی‌ها به عنوان نوار اطمینان پیش‌بینی منفرد انتخاب شد. این کار برای نقاط پرسش بعدی نیز تکرار شد. سپس برای کلیه نقاط پیش‌بینی، نوار اطمینان بصورت هاشور در شکل رسم شد. از روی شکل رسم شده نسبت به عدم قطعیت برای پیش‌بینی‌ها قضاوت شد طوری که اگر نوار اطمینان نزدیک به هم (باریک) می‌شد، عدم قطعیت ناچیز و در جایی که از هم باز می‌شد، عدم قطعیت بیشتر در نظر گرفته می‌شد.

با در نظر گرفتن روش جمعی، پیش‌بینی احتمالاتی (به ازای پارامترهای N_s که در بخش واسنجی انتخاب شده‌اند) انجام گرفت. مجموعه اعضای $\{\tilde{y}(t_0 + T)_k\}_s$ که در آن $(k = 1, \dots, N)$ در شکل یک توزیع تجربی کلی با اندازه $\sum_{s=1}^{N_s} r(s)$ با یکدیگر ادغام شدند. برای کاهش حجم محاسبات، یک تعداد مشخص $(r(s), s = 1, \dots, N_s)$ از اعضا به صورت تصادفی از هر یک از نمونه‌ها انتخاب شده و توزیع احتمالاتی کلی حاکم بر $\tilde{y}(t_0 + T)$ با مقادیر N_s ایجاد شد. میانه این توزیع به عنوان پیش‌بینی نقطه‌ای در نظر گرفته شد. روش پیش‌بینی احتمالاتی در ترکیب با روش پیش‌بینی غیرخطی این امکان را فراهم می‌سازد تا بتوان عدم قطعیت موجود در پیش‌بینی را محاسبه نمود.

روش جمعی احتمالاتی نزدیکترین همسایگی¹ (NNPE)

ابتدا روش K-NN معرفی و در ادامه به توصیف روش NNPE پرداخته می‌شود. روش K-NN جزء روش‌های رگرسیونی ناپارامتری است. یک تخمین‌گر رگرسیونی ناپارامتری، y^{NP} ، تابع هزینه زیر را با در نظر گرفتن یک سری زمانی به شکل $D_N = \{ \{y_i, x_i\} \in R_+^l \times R_+^l, i = 1, \dots, N \}$ روی D_N سری‌های زمانی D_N (داده‌های آموزشی) کمینه می‌کند:

$$C(y^{NP}) = \frac{\sum_{i=1}^N (y_i - y^{NP})^2 K(x_i, x_0, b)}{N} \quad (3)$$

در این رابطه، متغیر خروجی، با y_i و متغیرهای ورودی به صورت بردار x_i نشان داده می‌شوند. x_0 مشخص کننده نقطه پرسش² بوده و از مجموعه داده‌های آزمون در فضای ورودی انتخاب می‌شود. مجموعه دیگری تحت عنوان داده‌های آزمون یا D_M مشابه D_N در نظر گرفته شد، طوری که این داده‌ها هیچ عضو مشترکی با داده‌های مربوط به بخش آموزشی یا D_N نداشتند و نقطه پرسش نیز از بین این داده‌ها (مجموعه داده‌های آزمون) انتخاب و متناظر با متغیر مستقل مربوط به بردار x است. تابع K نیز معرف وزن یا تابع کرنل می‌باشد که مقدار آن با توجه به فاصله اقلیدسی هر نقطه در بخش آموزش از نقطه پرسش در بخش آزمون طبق رابطه 4 به دست می‌آید.

$$K(x_i, x_0, b) = \begin{cases} 1 & \text{if } \|x_i - x_0\| \leq b \\ 0 & \text{if } \|x_i - x_0\| > b \end{cases} \quad (4)$$

در این رابطه $\|\cdot\|$ نشان دهنده فاصله اقلیدسی و b شعاع همسایگی است. کمینه کردن معادله 3 با توجه به مقدار پارامتر y^{NP} و کرنل تعریف شده در رابطه 4 انجام می‌شود. برای تخمین y^m از

3- Calibration Set
4- Noncalibration Set

1- Nearest Neighbor Probabilistic Ensemble
2- Query Point

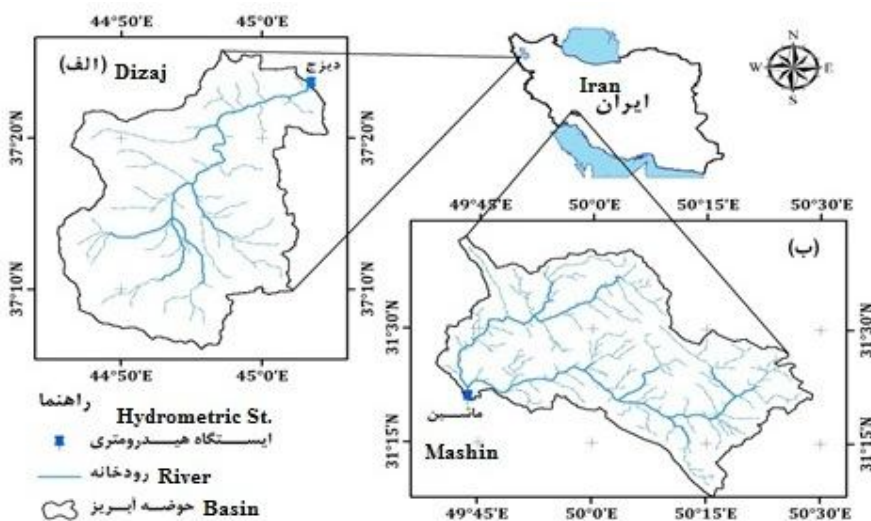
منطقه مورد مطالعه و داده‌ها

الگوهای مورد استفاده در این تحقیق در دو حوضه آبریز مختلف کشور به کار گرفته شدند. حوضه باراندوزچای با مساحت 1203 کیلومتر مربع، طول آبراهه اصلی برابر 75 کیلومتر و گستردگی جغرافیایی از ۴۵۴۴' تا ۱۴۴۵' طول شرقی و ۰۶۳۷' تا ۲۹۳۷' عرض شمالی در شمال غربی کشور واقع شده است. در این تحقیق از داده‌های جریان روزانه در ایستگاه دیزج واقع بر رودخانه باراندوزچای استفاده شد.

دیگر حوضه آبریز در این تحقیق حوضه آبریز زرد در شمال شهرستان رامهرمز و شرق استان خوزستان با وسعتی برابر با 875 کیلومتر مربع می‌باشد. این حوضه در مختصات ۳۹۴۹' تا ۱۰۵۰' طول شرقی و ۲۱۳۱' تا ۴۱۳۱' عرض شمالی واقع شده است. داده‌های

جریان روزانه در ایستگاه ماشین واقع در این حوضه استفاده شد. در شکل 1 موقعیت جغرافیایی حوضه‌ها و ایستگاه‌های هیدرومتری نشان داده شده است.

برای ایستگاه‌های آسنجی داده‌های جریان روزانه به مدت دوازده سال در طول سال‌های 1377 تا 1388 استفاده شدند. در طول این دوره، داده‌ها فاقد مفقودی بوده و از کیفیت آماری قابل قبولی برخوردار بودند. داده‌ها به سه بخش کلی شامل برآزش، واسنجی و آزمون مدل تقسیم‌بندی شدند. شش سال اول داده‌ها جهت برآزش، سه‌سال بعدی برای واسنجی و 3 سال پایانی جهت آزمون الگوها استفاده شدند. پارامترهای آماری داده‌های جریان در ایستگاه‌های دیزج و ماشین در جدول 1 نشان داده شده است.



شکل 1- موقعیت جغرافیایی منطقه مورد مطالعه (الف - حوضه آبریز باراندوزچای - حوضه آبریز زرد) Figure 1- The geographic location of study area (a- Baranduz-Chay basin b- Zard basin)

جدول 1- برخی از مهمترین مشخصه های آماری داده‌های دبی جریان روزانه رودخانه‌های مورد مطالعه Table 1- Some of important statistical characteristics of daily discharge data for studied rivers

نام رودخانه (River)	مساحت حوضه (Basin area) (km ²)	میانگین (mean) (m ³ /s)	حداکثر (maximum) (m ³ /s)	حداقل (minimum) (m ³ /s)	انحراف معیار (standard deviation) (m ³ /s)	ضریب تغییرات (coefficient of variation)	ضریب چولگی (skewness) (m ³ /s)
باراندوزچای (Baranduz-Chay)	1203	5.68	102	0	8.36	1.47	3.90
رود زرد (Zard River)	875	7.28	295	0	15.12	2.08	10.00

مربعات خطا (RMSE) و ضریب کارایی ناش - ساتکلیف (E) مطابق روابط زیر استفاده شد.

شاخص‌های خطا سنجی مدل

برای ارزیابی عملکرد الگوهای به کار گرفته شده در این تحقیق از چندین شاخص آماری شامل ضریب همبستگی (R)، ریشه میانگین

است. با توجه به جدول 2، می‌توان گفت که روش جمعی احتمالاتی غیرخطی عملکرد مناسبی در هر دو حوضه داشته است. قابل ذکر است که نتایج ارائه شده در جدول 2، به ازای میانه پیش‌بینی‌های جمعی ارائه شده است. با توجه به جدول 2 می‌توان دید که مقادیر R در ایستگاه‌های مختلف معنی‌دار است. در شکل‌های 2 و 3 روند الگوسازی جریان با روش جمعی احتمالاتی غیرخطی برای یک دوره 200 (1 دی 87 تا 17 تیر 88) روزه از مجموعه داده‌های آزمون به ترتیب برای ایستگاه‌های دیزج و ماشین ارائه شده است. در این شکل‌ها باند اطمینان 95 درصد که نشان دهنده عدم قطعیت ناشی از پیش‌بینی انجام شده توسط مدل می‌باشد، نشان داده است.

با توجه به شکل‌های 2 و 3 می‌توان دید که با افزایش مقادیر دبی پهنای نوار اطمینان نیز افزایش یافته است به طوری که بیشترین پهنای نوار به مقادیر دبی حداکثر اختصاص دارد. بنظر می‌رسد، این مساله حاکی از عدم قطعیت بیشتر در پیش‌بینی مقادیر حداکثر می‌باشد. چنین عملکردی نشان می‌دهد که دقت الگو در پیش‌بینی مقادیر حداقل بهتر از مقادیر حداکثر بوده است. این مطلب در مطالعات تام و همکاران (22) نیز اشاره شده است. اگر چهار یک ارزیابی کلی می‌توان گفت که عملکرد الگو در پیش‌بینی مقادیر حداکثر دبی قابل قبول بوده است و تنها تعداد معدودی از نقاط مشاهداتی در خارج از باند اطمینان قرار دارند. همچنین عملکرد الگو در پیش‌بینی مقادیر دبی جریان در ایستگاه ماشین در مقایسه با ایستگاه دیزج بهتر بوده است که این مساله نتایج ارائه شده در جدول 2 را تأیید می‌کند. دلیل این موضوع احتمالاً می‌تواند به وسعت کمتر حوضه روزد زرد در مقایسه با حوضه باراندوزچای نسبت داده شود. نتایج ارائه شده در جدول 2 و شکل‌های 2 و 3 با مطالعات تام و همکاران (22) مطابقت دارد. در مطالعات تاما نیز به ازای داده‌های مورد استفاده در دوره آزمون مقادیر به دست آمده برای شاخص‌های عملکرد R و E مشابهت نزدیکی با مقادیر آن‌ها در این تحقیق دارد.

$$R = \sqrt{1 - \frac{\sum_{i=1}^N (O_i - S_i)^2}{\sum_{i=1}^N O_i^2 - \frac{\sum_{i=1}^N S_i^2}{N}}} \quad (7)$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (O_i - S_i)^2} \quad (8)$$

$$E = 1 - \frac{\sum_{i=1}^N (O_i - S_i)^2}{\sum_{i=1}^N (O_i - \bar{O}_i)^2} \quad (9)$$

در روابط فوق، N ، تعداد داده‌ها، مقدار O_i ، مقدار مشاهداتی برای جریان رودخانه و S_i ، مقدار پیش‌بینی شده برای جریان رودخانه توسط مدل و \bar{O}_i میانگین مقادیر مشاهداتی می‌باشد.

نتایج و بحث

در بخش اول نتایج روش پیش‌بینی جمعی احتمالاتی غیرخطی برای حوضه‌های باراندوزچای و رود زرد ارائه می‌شود. به طوریکه اشاره شد، برای این روش ابتدا با تعدادی آزمون اولیه بر روی داده‌های واسنجی مقادیر پارامترها مشخص شدند. با توجه به اینکه با در نظر گرفتن مقادیر کوچک برای r به دلیل وجود اغتشاشات و نویزها نتایج مناسبی حاصل نمی‌شود، لذا حداقل مقدار برای پارامتر r برابر 100 در نظر گرفته شد و مقدار آن به صورت مضربی از 100 شامل 200، 300، 500، 1000، 2000، 5000، 10000 افزایش یافت. مقدار پارامتر m نیز در محدوده 1 تا 20 در نظر گرفته شد. پارامتر تأخیر زمانی γ نیز در محدوده 1 تا 5 بررسی شد. در ادامه کار دقت پیش‌بینی در مجموعه داده‌های واسنجی با استفاده از شاخص $RMSE$ ارزیابی شد. در هر دو حوضه، تعداد عضوهای پارامترهای جمعی (N_s) برابر 100 که به ازای بهترین عملکرد در مجموعه داده‌های واسنجی حاصل شد، در نظر گرفته شد.

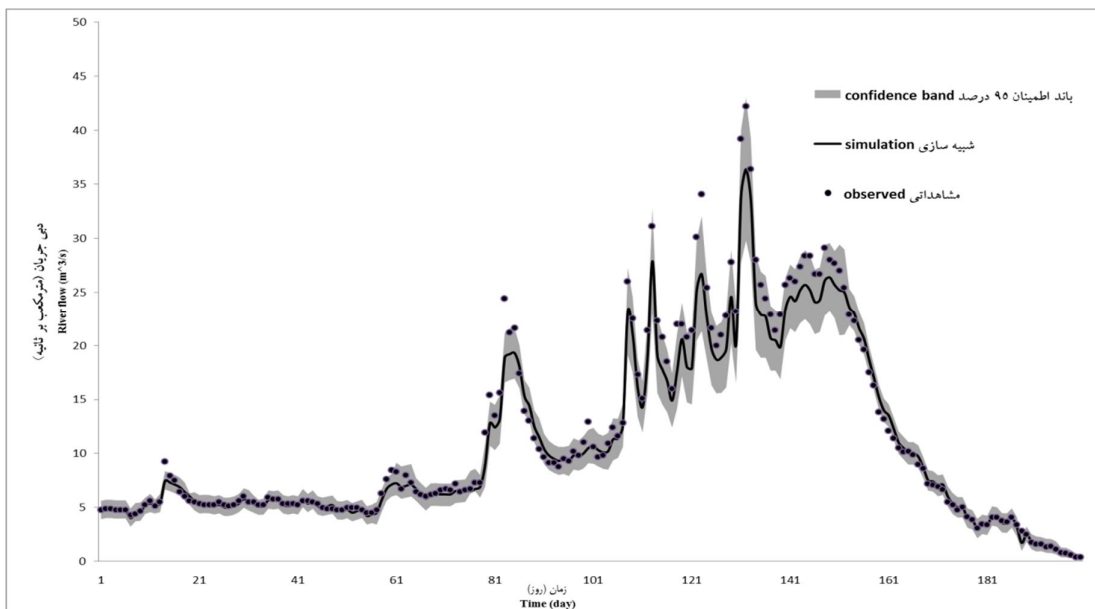
در جدول 2، شاخص‌های آماری روش جمعی احتمالاتی غیرخطی به ازای مجموعه داده‌های آزمون در ایستگاه‌های مختلف ارائه شده

جدول 2- شاخص‌های آماری عملکرد روش جمعی احتمالاتی غیرخطی در پیش‌بینی جریان رودخانه های مورد مطالعه در مرحله آزمون

Table 2- The statistical performance criteria for nonlinear probabilistic ensemble method in discharge prediction of studied rivers during test period

ایستگاه (Station)	R	RMSE (m ³ /s)	E
دیزج (Dizaj)	0.92*	4.64	0.88
ماشین (Mashin)	0.94*	2.49	0.91

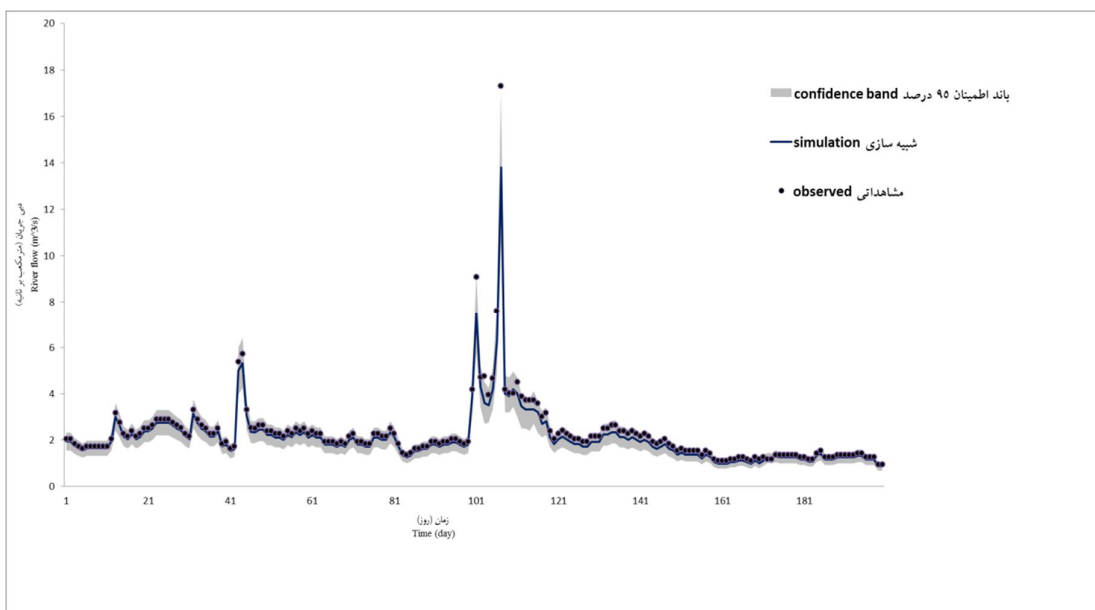
علامت * نشان دهنده معنی‌دار بودن مقدار R می‌باشد.
* indicated that R is significant in 5% level.



شکل 2- عملکرد روش پیش‌بینی جمعی احتمالاتی غیرخطی در مدل‌سازی دبی جریان ایستگاه دیزج در حوضه رودخانه باراندوزچای نواحی هاشور خورده باند اطمینان 95 درصد برای پیش‌بینی هستند.

Figure 2- The Performance of nonlinear probabilistic ensemble method in river flow prediction of Baranduz-Chay basin (Dizaj station)

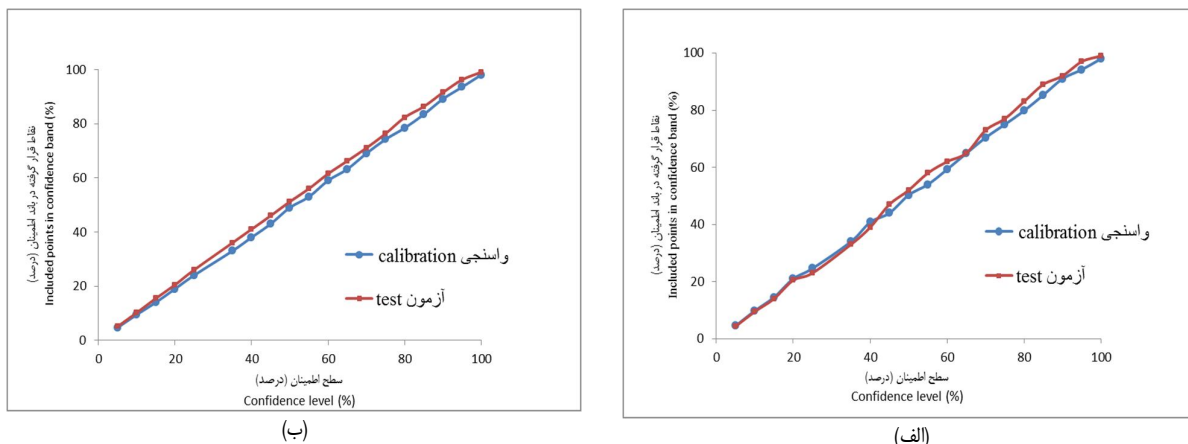
The shaded area represent the 95% confidence interval of prediction



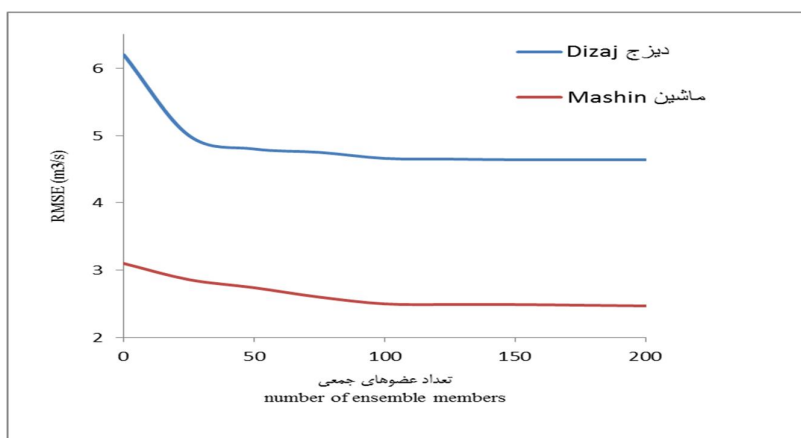
شکل 3- عملکرد روش پیش‌بینی جمعی احتمالاتی غیرخطی در مدل‌سازی دبی جریان ایستگاه ماشین در حوضه رودخانه زرد نواحی هاشور خورده باند اطمینان 95 درصد برای پیش‌بینی هستند.

Figure 3- The Performance of nonlinear probabilistic ensemble method in river flow prediction of Zard basin (Mashin station)

The shaded area represent the 95% confidence interval of prediction



شکل 4- درصد نقاط قرار گرفته در باند اطمینان روش جمعی احتمالاتی غیرخطی در مراحل واسنجی و آزمون مدل الف) دیزج ب) ماشین
Figure 4- Percentage of values included in the confidence band of the probabilistic NLP during calibration and test step a) Dizaj b) Mashin



شکل 5- روند تغییرات شاخص RMSE نسبت تعداد عضوهای جمعی در ایستگاههای مختلف
Figure 5- the variation of RMSE respect to the number of ensemble members for different stations

استفاده کردند، مطابقت دارد. این نتایج نشان می‌دهد که یک روش پیش‌بینی جمعی شامل مجموعه‌ای از پیش‌بینی‌های منفرد عملکرد بهتری در مقایسه با پیش‌بینی منفرد دارد. عملکرد بهتر روش جمعی به این دلیل است که در آن از مجموعه پارامترهای بهینه در ساختار الگو استفاده می‌گردد و این نکته می‌تواند موجب بهبود عملکرد روش جمعی در مقایسه پیش‌بینی منفرد باشد.

در شکل 5 نیز تغییرات شاخص خطای $RMSE$ به ازای تعداد عضوهای جمعی در ایستگاه‌های مختلف ارائه شده است. با توجه به این شکل می‌توان دید که مقدار شاخص $RMSE$ با افزایش تعداد عضوهای جمعی کاهش یافته است. برای ایستگاه دیزج به ازای تعداد عضوهای در محدوده 50 تا 100 این تغییرات روند خطی داشته است. تغییرات $RMSE$ در ایستگاه ماشین نیز به ازای تعداد عضوهای جمعی در محدوده 100 تا 150 تقریباً ثابت است. به طور کلی می‌توان گفت

تعداد نقاط واقع شده در نوار اطمینان در مقایسه با سطح اطمینان ترسیم شده می‌تواند شاخص مهمی جهت قضاوت در خصوص دقت محدوده اطمینان باشد. در شکل 4 نتایج این تحلیل برای مجموعه داده‌های بخش واسنجی و آزمون در ایستگاه‌های دیزج و ماشین نشان داده شده است. تطابق بسیار خوبی بین توزیع احتمالاتی نظری و تجربی به ازای داده‌های واسنجی و آزمون برای هر دو ایستگاه وجود دارد و روند خطی موجود کاملاً مشهود است. این مساله حاکی از آن است که رفتار پویای موجود در سری زمانی داده‌های مورد بررسی بخوبی توسط الگوارائه شده است. با توجه به اینکه در این تحقیق از یک دوره آماری 12 ساله جهت الگوسازی استفاده شده است، به نظر می‌رسد با افزایش طول دوره آماری نتایج نیز بهبود یابد. نتایج ارائه شده در این بخش با تحقیقات ریگوندا و همکاران (17) که از چند جمله‌ای‌های ناحیه‌ای برای برآورد سری‌های زمانی به صورت جمعی

غیرخطی) برای ایستگاه‌های دیزج و ماشین ارائه شده است. با توجه به این شکل‌ها می‌توان دید که عملکرد الگوی NNPE در پیش‌بینی مقادیر جریان به ازای هر دو ایستگاه مناسب و قابل قبول بوده است. در این روش نیز همانند روش جمعی احتمالاتی غیرخطی با افزایش مقدار دبی پهنای نوار اطمینان افزایش یافته است. عملکرد الگو در پیش‌بینی مقادیر حداکثر جریان مناسب بوده است. با بررسی نتایج به دست آمده مشخص گردید که برای داده‌های دیزج و ماشین به ترتیب تعداد 108 و 96 نقطه از مجموعه نقاط داده‌های آزمون (1095 نقطه) خارج از محدوده نوار اطمینان قرار دارند که این مساله نشان از دقت قابل قبول روش NNPE در پیش‌بینی جریان می‌باشد.

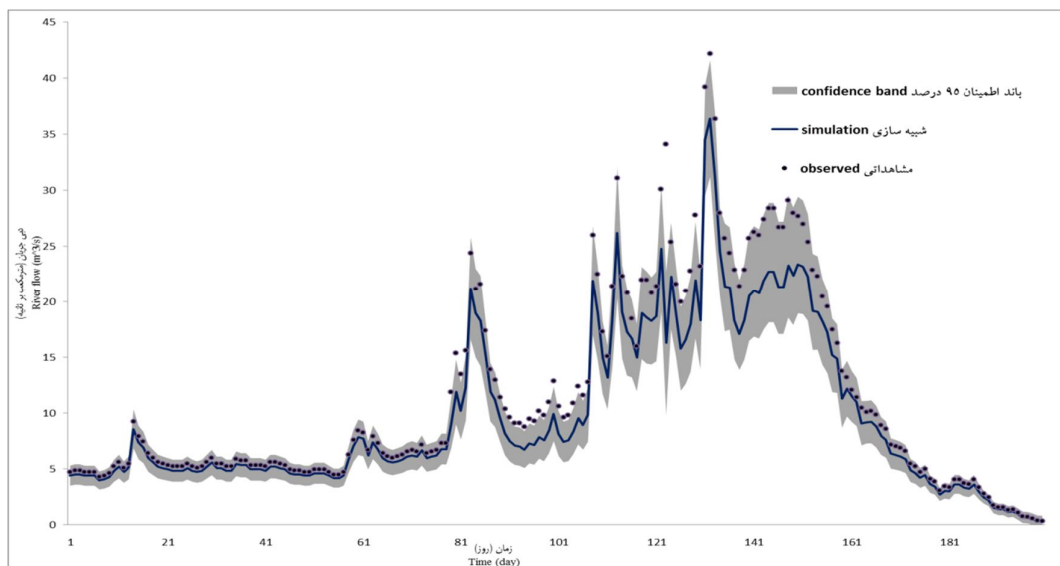
در هر دو ایستگاه، به ازای تعداد عضوهای جمعی برابر 100 روند تغییرات RMSE تقریباً ناچیز است و این تعداد جهت الگوسازی مناسب به نظر می‌رسد.

در ادامه به بررسی عملکرد روش NNPE در پیش‌بینی جریان رودخانه در ایستگاه‌های دیزج و ماشین پرداخته می‌شود. در جدول 3 شاخص‌های عملکرد الگو در پیش‌بینی جریان به ازای داده‌های دوره آزمون ارائه شده است. با توجه به مقادیر جدول می‌توان گفت که عملکرد الگو در ایستگاه ماشین در مقایسه با ایستگاه دیزج نسبتاً بهتر بوده است. در شکل‌های 6 و 7 شبیه‌سازی انجام شده با الگوی NNPE به ازای یک دوره 200 روزه از مجموعه داده‌های بخش آزمون (مشابه دوره انتخابی برای روش جمعی احتمالاتی

جدول 3- شاخص‌های آماری عملکرد روش NNPE در پیش‌بینی جریان رودخانه در مرحله آزمون
Table 3- The statistical performance criteria for NNPE model in river flow prediction for test period

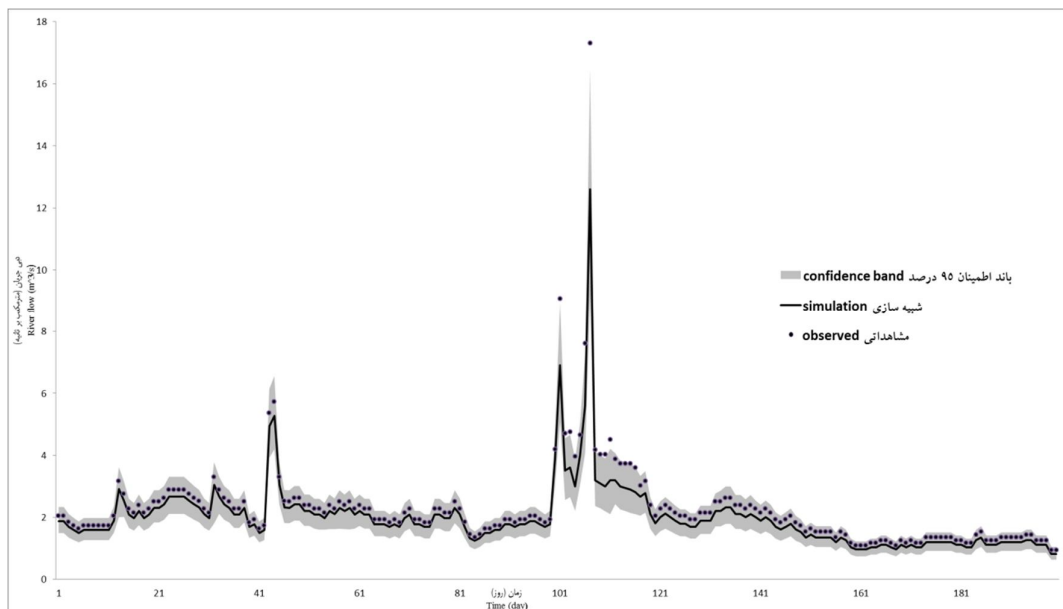
ایستگاه (Station)	R	RMSE (m ³ /s)	E
دیزج (Dizaj)	0.95*	4.17	0.91
ماشین (Mashin)	0.97*	2.01	0.93

علامت * نشان دهنده معنی‌دار بودن مقدار R در سطح 5 درصد می‌باشد.
* indicated that R is significant in 5% level.



شکل 6- عملکرد روش NNPE در مدل‌سازی دبی جریان ایستگاه دیزج در حوضه باراندوزچای
نواحی هاشور خورده باند اطمینان 95 درصد برای پیش‌بینی هستند.

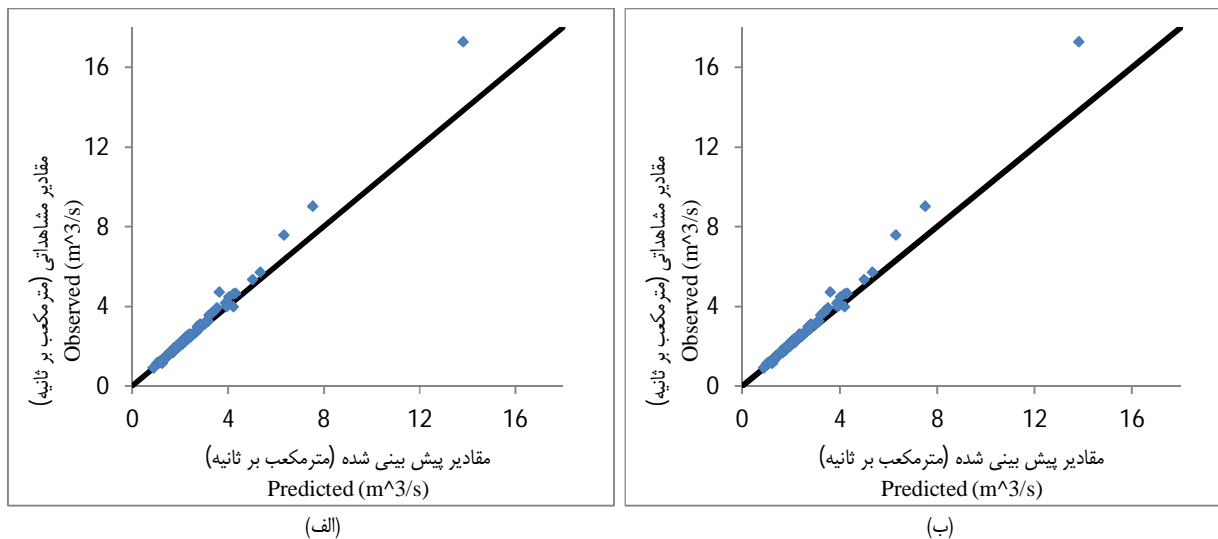
Figure 7- The Performance of NNPE model in river flow prediction of Baranduz-Chay basin (Dizaj station)
The shaded area represent the 95% confidence interval of prediction



شکل 7- عملکرد روش NNPE در مدل‌سازی دبی جریان ایستگاه ماشین در حوضه رود زرد

نواحی هاشور خورده باند اطمینان 95 درصد برای پیش بینی هستند.

Figure 7- The Performance of NNPE model in river flow prediction of Zard basin (Mashin station)
The shaded area represent the 95% confidence interval of prediction



شکل 8- نمودار پراکندگی داده‌های دبی جریان برای روش NNPE الف) دیزج ب) ماشین
Figure 8- Scatterplots of river flow data for NNPE model a) Dizaj b) Mashin

مشاهداتی به خود می‌گیرد. به عبارت دیگر نقاط بیشتر در دبی‌های کوچک به خط نیمساز نزدیک‌ترند و در دبی‌های بالا از خط یک به یک فاصله گرفته‌اند. چنین رفتاری در نمودار پراکندگی داده‌ها منطقی بنظر می‌رسد و همان گونه که قبلاً نیز ذکر شد نشان می‌دهد دقت الگو در شبیه‌سازی مقادیر حداقل جریان در مقایسه مقادیر حداکثر

در شکل 8 نیز نمودار پراکندگی داده‌های شبیه‌سازی شده در مقایسه با مقادیر مشاهداتی برای ایستگاه‌های دیزج و ماشین ارائه شده است. این شکل نتایج ارائه شده در شکل‌های 6 و 7 را تأیید می‌کند. با توجه به این شکل می‌توان نتیجه گرفت که هرچه دبی جریان بیشتر می‌شود مقدار پیش‌بینی انحراف بیشتری از مقدار

بهتر می‌باشد.

با توجه به اینکه در روش NNPE پیش‌بینی‌ها بر مبنای الگوریتم نزدیکترین همسایگی صورت می‌گیرد لذا لازم است مقدار پارامترهای شعاع همسایگی (b) و تعداد تاخیرها (l) در پیش‌بینی‌های انجام شده مورد توجه قرار گیرد. به طور کلی با افزایش مقدار دبی مقدار این پارامترها نیز افزایش یافت. مقادیر شعاع همسایگی و تعداد تاخیرها به ازای حداکثر مقدار دبی در ایستگاه دیزج به ترتیب برابر 0/45 و 5 به دست آمدند. نتایج به دست آمده در این بخش با مطالعات پانده و همکاران (14) مطابقت دارد. در مطالعات آن‌ها نیز با افزایش مقدار دبی پهنای نوار اطمینان افزایش یافته است.

مقایسه کلی نتایج به دست‌آمده با روش NNPE و روش جمعی احتمالاتی غیرخطی حاکی از برتری نسبی روش NNPE در پیش‌بینی‌ها دارد. عرض پهنای نوار به دست آمده برای روش NNPE در مقایسه با روش جمعی احتمالاتی غیرخطی کمتر می‌باشد که ناشی از عدم قطعیت کمتر روش NNPE و دقت بیشتر آن می‌باشد. چنین مسأله‌ای می‌تواند ناشی از عملکرد بهتر الگوریتم نزدیکترین همسایگی در ساختار روش NNPE در مقایسه با عملگرهای خطی ناحیه‌ای در ساختار روش جمعی احتمالاتی غیرخطی باشد. همچنین نتایج به دست آمده نشان می‌دهد که عملکرد هر دو روش به ازای داده‌های ایستگاه ماشین مناسب‌تر بوده است. این نتیجه می‌تواند ناشی از وسعت کمتر حوضه رود زرد در مقایسه حوضه باراندوزچای باشد. همچنین دامنه تغییرات مقادیر دبی جریان در ایستگاه ماشین در مقایسه با ایستگاه دیزج به ازای مجموعه داده‌های آزمون الگوها کمتر بوده است که به نظر می‌رسد این مسأله می‌تواند بر دقت الگوسازی موثر باشد. در این تحقیق فقط از داده‌های رواناب جهت الگوسازی استفاده شده است. استفاده از سری داده‌های بارش و دما مخصوصاً در حوضه‌های برفی ممکن است موجب افزایش دقت الگوشود و عدم قطعیت مدل را کاهش دهد به عبارت دیگر نوار اطمینان الگو هر چه بیشتر باریکتر شود که می‌تواند موضوع تحقیقات آتی باشد. افزون بر این در مطالعه حاضر از داده‌های دبی جریان در مقیاس زمانی روزانه استفاده شد، به نظر می‌رسد دقت پیش‌بینی با داده‌های دبی جریان در مقیاس زمانی ساعتی در افزایش دقت الگو بسیار مفید باشد. در نمودارهای ارائه شده برای دوره 200 روزه با الگوی NNPE، مقدار دبی حداکثر در ایستگاه‌های دیزج و ماشین به ترتیب به میزان 14 و

20 درصد نسبت به مقادیر مشاهداتی کمتر برآورد شده اند. همچنین در این دوره 200 روزه، مقدار حجم رواناب پیش‌بینی شده برای حوضه‌های باراندوزچای و رود زرد به ترتیب به میزان 16 و 9 درصد نسبت به مقادیر مشاهداتی کمتر برآورد شده‌اند.

نتیجه‌گیری کلی

در این مطالعه به عملکرد روش‌های جمعی غیرخطی احتمالاتی و جمعی احتمالاتی مبتنی بر الگوریتم نزدیکترین همسایگی در پیش‌بینی جریان رودخانه پرداخته شد. از داده‌های دبی روزانه رودخانه باران دوزچای در ایستگاه دیزج در شمال غرب و رود زرد در ایستگاه ماشین در جنوب غرب کشور استفاده شد. از الگوهای ورودی مختلف شامل مقادیر دبی‌های پیشین جهت الگوسازی دبی استفاده شد. نتایج به دست آمده حاکی از عملکرد مناسب و دقت بالای روش‌ها در پیش‌بینی مقادیر جریان روزانه بود. یکی از ضعف‌های عمده الگوهای هیدرولوژیکی در پیش‌بینی مقادیر دبی اوج جریان می‌باشد که در این خصوص هر دو الگو عملکرد قابل قبولی داشتند. به کارگیری الگوهای استفاده شده در این تحقیق در قالب شکل جمعی امکان بررسی کمی عدم قطعیت ناشی از پیش‌بینی جریان را فراهم نمود که این مسأله جهت بکارگیری الگو برای پیش‌بینی در اقل‌های آبی اهمیت فراوانی دارد. مقایسه عملکرد الگوهای مختلف در این مطالعه حاکی از عملکرد بهتر و قطعیت بیشتر روش جمعی احتمالاتی نزدیکترین همسایگی در مقایسه با روش جمعی احتمالاتی غیرخطی در پیش‌بینی مقادیر دبی جریان بود. همچنین بررسی عملکرد الگوها در حوضه‌های مختلف نیز نشان داد که الگوهای استفاده شده در حوضه آبریز رودخانه زرد عملکرد مناسب‌تری در مقایسه با حوضه آبریز باراندوزچای داشته‌اند که این مسأله می‌تواند ناشی از وسعت کمتر حوضه رودخانه زرد باشد. همچنین در این حوضه نوسانات و دامنه تغییرات مقادیر دبی جریان در دوره آزمون الگوها نسبت به حوضه باراندوزچای کمتر بود. به طور کلی می‌توان نتیجه گرفت که استفاده از روش‌های پیش‌بینی جمعی در مقایسه با پیش‌بینی‌های منفرد تصمیم‌گیران صنعت آب را در لحاظ نمودن عدم قطعیت در پیش‌بینی یاری می‌کند. استفاده از این روش‌ها در پیش‌بینی جریان سایر رودخانه‌ها توصیه می‌شود.

منابع

- 1- Azmi M., and Araghinejad S. 2012. Developed K-nearest neighbor method for river flow prediction. Journal of Water and Wastewater, 2: 108-119.
- 2- Alfieri L., Thielen J., and Pappenberger F. 2012. Ensemble hydro-meteorological simulation for flash flood early detection in southern Switzerland. Journal of Hydrology, 424: 143-153.
- 3- Casdagli M. 1992. Chaos and deterministic versus stochastic nonlinear modeling. Journal of the Royal Statistical Society, Series B (Methodological), 54(2): 303-328.

- 4- Chatfield C. 2001. Prediction intervals, in Principles of Forecasting: A Handbook for Researchers and Practitioners, edited by J. Armstrong, Springer, New York.
- 5- Fan F. M., Collischonn W., Meller A., and Botelho L.C.M. 2014. Ensemble streamflow forecasting experiments in a tropical basin: The Sao Francisco river case study. *Journal of Hydrology*, In Press.
- 6- Hampel F.R. 1974. The influence curve and its role in robust estimation. *Journal of the American Statistical Association*, 346: 383-393.
- 7- Herr H.D., and Krzysztofowicz R. 2010. Bayesian ensemble forecast of river stages and ensemble size requirements. *Journal of Hydrology*, 387: 151-164.
- 8- Kantz H., and Schreiber T. 1997. *Nonlinear Time Series Analysis*. Cambridge Univ. Press, New York.
- 9- Kember G., Flower A.C., and Holubeshen J. 1993. Forecasting river flow using nonlinear dynamics. *Stochastic Hydrology and Hydraulics*, 7: 205-212.
- 10- Krzysztofowicz R. 1999. Bayesian theory of probabilistic forecasting via deterministic hydrologic model, *Water Resources Research*, 35(9), 2739-2750.
- 11- Kuczera G., and Parent E. 1998. Monte Carlo assessment of parameter uncertainty in conceptual catchment models: The Metropolis algorithm, *Journal of Hydrology*, 211: 69-85.
- 12- Laio F., Porporato A., Revelli R., and Ridolfi L. 2003. A comparison of nonlinear flood forecasting methods, *Water Resources Research*, 39(5).
- 13- Montanari A., and Brath A. 2004. A stochastic approach for assessing the uncertainty of rainfall-runoff simulations. *Water Resources Research*, 40(1).
- 14- Pande P., McKee M., and Bastidas L.A. 2009. Complexity-based robust hydrologic prediction. *Water Resources Research*, 45, W10406.
- 15- Phoon K. K., Islam M. N., Liaw C. Y., and Liong S.Y. 2002. Practical inverse approach for forecasting nonlinear hydrological time series. *Journal of Hydrologic Engineering*, 7(2): 116-128.
- 16- Porporato A., and Ridolfi L. 1997. Nonlinear analysis of river flow time sequences. *Water Resources Research*, 33(6): 1353-1367.
- 17- Regonda S.K., Rajagopalan B., Lall M., Clark U., and Moon Y.I. 2005. Local polynomial method for ensemble forecast of time series. *Nonlinear Processes Geophysics*, 12(3): 397-406.
- 18- Regonda S.K., Rajagopalan B., Clark M., and Zagona E. 2006. A multimodel ensemble forecast framework: Application to spring seasonal flows in the Gunnison River Basin. *Water Resources Research*, 42, W09404.
- 19- Silville J.K., Ahlquist J.E., and Toth Z. 1997. An ensemble forecasting primer. *Weather and Forecasting*, 12: 809-818.
- 20- Sivakumar B. 2000. Chaos theory in hydrology: Important issues and interpretations. *Journal of Hydrology*, 227: 1-20.
- 21- Sharma A., and Lall U. 1999. A nonparametric approach for daily rainfall simulation. *Mathematics and Computers in Simulation*, 48: 361-371.
- 22- Tamea S., Laio F., and Ridolfi L. 2005. Probabilistic nonlinear prediction of river flows. *Water Resources Research*, 41, W09421.
- 23- Todini E. 2004. Role and treatment of uncertainty in real-time flood forecasting. *Hydrological Processes*, 18(14): 2743-2746.
- 24- Wu C.L., and Chau K.W. 2010. Data-driven models for monthly streamflow time series prediction. *Engineering Applications of Artificial Intelligence*, 23: 1350-1367.

River Flow Prediction Using the Nearest Neighbor Probabilistic Ensemble Method

H. Sanikhani^{5*} - Y. Dinpashoh² - M. A. Ghorbani³ - M. Zarghami⁴

Received: 07-06-2014

Accepted: 27-10-2014

Introduction: In the recent years, researchers interested on probabilistic forecasting of hydrologic variables such river flow. A probabilistic approach aims at quantifying the prediction reliability through a probability distribution function or a prediction interval for the unknown future value. The evaluation of the uncertainty associated to the forecast is seen as a fundamental information, not only to correctly assess the prediction, but also to compare forecasts from different methods and to evaluate actions and decisions conditionally on the expected values. Several probabilistic approaches have been proposed in the literature, including (1) methods that use resampling techniques to assess parameter and model uncertainty, such as the Metropolis algorithm or the Generalized Likelihood Uncertainty Estimation (GLUE) methodology for an application to runoff prediction), (2) methods based on processing the forecast errors of past data to produce the probability distributions of future values and (3) methods that evaluate how the uncertainty propagates from the rainfall forecast to the river discharge prediction, as the Bayesian forecasting system.

Materials and Methods: In this study, two different probabilistic methods are used for river flow prediction. Then the uncertainty related to the forecast is quantified. One approach is based on linear predictors and in the other, nearest neighbor was used. The nonlinear probabilistic ensemble can be used for nonlinear time series analysis using locally linear predictors, while NNPE utilize a method adapted for one step ahead nearest neighbor methods. In this regard, daily river discharge (twelve years) of Dizaj and Mashin Stations on Baranduz-Chay basin in west Azerbaijan and Zard-River basin in Khouzestan provinces were used, respectively. The first six years of data was applied for fitting the model. The next three years was used to calibration and the remained three years utilized for testing the models. Different combinations of recorded data were used as the input pattern to streamflow forecasting.

Results and Discussion: Application of the used approaches in ensemble form (in order to choice the optimized parameters) improved the model accuracy and robustness in prediction. Different statistical criteria including correlation coefficient (R), root mean squared error (RMSE) and Nash-Sutcliffe efficiency coefficient (E) were used for evaluating the performance of models. The ranges of parameter values to be covered in the ensemble prediction have been identified by some preliminary tests on the calibration set. Since very small values of k have been found to produce unacceptable results due to the presence of noise, the minimum value is fixed at 100 and trial values are taken up to 10000 (k = 100, 200, 300, 500, 1000, 2000, 5000, 10000). The values of γ were chosen between 1 and 20 and delay time values γ are tested in the range [1,5]. With increasing the discharge values, the width of confidence band increased and the maximum confidence band is related to maximum river flows. In Dizaj station, for ensemble numbers in the range of 50-100, the variation of RMSE is linear. The variation of RMSE in Mashin station is linear for ensemble members in the range of 100-150. It seems the numbers of ensemble members equals to 100 is suitable for pattern construction. The performance of NNPE model was acceptable for two stations. The number of points excluded 95% confidence interval were equal to 108 and 96 for Dizaj and Mashin stations, respectively. The results showed that the performance of model was better in prediction of minimum and median discharge in comparing maximum values.

Conclusion: The results confirmed the performance and reliability of applied methods. The results indicated the better performance and lower uncertainty of ensemble method based on nearest neighbor in comparison with probabilistic nonlinear ensemble method. Nash-Sutcliffe model efficiency coefficient (E) for nearest neighbor probabilistic ensemble method in Dizaj and Mashin Stations during test period of model obtained 0.91 and 0.93, respectively. The investigation on the performance of models in different basins showed that the models have better performance in Zard river basin compared to Baranduz-Chaybasin. Furthermore the variation of discharge

1, 2, 3- PhD Student and Associate Professors Water Resources Engineering Department, Agriculture Faculty, University of Tabriz, Respectively

(*-Corresponding Author Email: hsanikhani12@gmail.com)

4- Associate Professor, Water Engineering Department, Civil Engineering Faculty, University of Tabriz

values during test period in Zard basin was lower in comparison of Baranduz-Chay basin. The real advantage of including streamflow forecasts requires detailed and specific investigations, but the preliminary results suggest the good potentiality of probabilistic NLP method. Using ensemble prediction method can help to decision makers in order to determine the uncertainty of prediction in water resources field.

Keywords: Baranduz-Chay, Dizaj, Hydrologic Variables, Zard River, Mashin