

## بررسی کارایی روش‌های ناپارامتریک مبتنی بر تجزیه و تحلیل باقیمانده‌ها و پارامتریک در برآورد عدم قطعیت

### مدل هیدرولوژیکی

ابوالحسن فتح‌آبادی<sup>۱\*</sup>، حامد روحانی<sup>۲</sup>، سید مرتضی سیدیان<sup>۳</sup>

۱. استادیار، گروه مرتع و آبخیزداری، دانشگاه گنبد کاووس، استان گلستان.

۲. استادیار، گروه مرتع و آبخیزداری، دانشگاه گنبد کاووس، استان گلستان.

۳. استادیار، گروه مرتع و آبخیزداری، دانشگاه گنبد کاووس، استان گلستان.

(تاریخ دریافت: ۱۳۹۵/۱۱/۱۸ - تاریخ بازنگری: ۱۳۹۶/۶/۲۲ - تاریخ تصویب: ۱۳۹۶/۶/۲۵)

### چکیده

به منظور تصمیم‌گیری مناسب جهت اجرای اقدامات مدیریتی نیاز است تا علاوه بر خروجی مدل دامنه عدم قطعیت آن نیز برآورد گردد. در تحقیق حاضر کارایی روش‌های ناپارامتریک (LEC (Local Errors and Clustering، رگرسیون چندک و جنگل تصادفی در برآورد عدم قطعیت مدل یکپارچه HBV در حوضه چهل‌چای استان گلستان بررسی گردید. پس از بهینه‌سازی پارامترهای مدل HBV با استفاده از روش تکامل تصادفی جوامع، مدل برای دوره‌های واسنجی و صحت سنجی اجرا و مقادیر باقیمانده‌ها محاسبه گردید. نتایج نشان داد با در نظر گرفتن متغیرهای دبی برآوردی، دبی مشاهداتی، مقدار بارش و مقادیر باقیمانده‌ها در حوضه مورد مطالعه داده‌های ورودی در چهار خوشه فازی قرار می‌گیرند. نتایج برآورد عدم قطعیت نشان داد بزرگترین و کوچکترین مقدار دامنه عدم قطعیت به ترتیب توسط روش‌های LEC در حالتی که توسط ماشین بردار رگرسیون آموزش دیده باشد و روش جنگل تصادفی، بدست آمده است. با توجه به مقادیر شاخص‌های ارزیابی (PICP (Prediction Interval Coverage Probability، MPI (Mean Prediction Interval) و ARIL (Average Relative Interval Length) بهترین عملکرد مربوط به روش رگرسیون چندک و سپس روش LEC در حالتی که آموزش داده نشده است، بود. در مقایسه با روش‌های ناپارامتریک، روش (Generalized Likelihood Uncertainty) (GLUE Estimation) با توجه به مقادیر هر سه معیار ارزیابی عملکرد مناسبی نداشت.

واژه‌های کلیدی: بارش رواناب، جنگل تصادفی، رگرسیون چندک، GLUE

### مقدمه

قطعیت آن نیز ارائه گردد. Chang *et al.* (1993) بیان می‌کنند عدم قطعیت جزئی از سیستم‌های طبیعی است و نمی‌توان آن را نادیده گرفت. در هیدرولوژی عدم قطعیت به مفهوم یک شاخص کمی از قابلیت اطمینان یک کمیت هیدرولوژی اندازه‌گیری شده یا خروجی یک مدل هیدرولوژیکی می‌باشد. این شاخص ممکن است به صورت خطای مدل یا یک دامنه‌ای از تغییرات کمیت مورد نظر باشد (Montanari, 2011). عدم قطعیت به دو گروه تصادفی (Aleatory); که منعکس کننده تغییرپذیری یک پدیده تکرار پذیر است و ریشه‌ای (Epistemic) که ناشی از نقص درک انسان از فرآیند است و با افزایش دانش فرد این نوع عدم قطعیت کاهش می‌یابد، تقسیم می‌شود (Walker *et al.*, 2003; Matott *et al.*, 2009).

با توجه به اهمیت عدم قطعیت در مدل‌سازی محققین مختلف روش‌های متفاوتی را برای برآورد عدم قطعیت ارائه داده‌اند که شامل شش دسته کلی ۱- روش‌های تحلیلی ۲-

به دلیل پیچیدگی فرآیندها و مؤلفه‌های مرتبط با چرخه آب در طبیعت، جهت شناخت بهتر ماهیت غیرخطی و پیچیده آن نیاز به توسعه مدل‌های شبیه‌سازی ساده شده با توجه به اهداف مورد انتظار می‌باشد. از این نظر تلخیص شامل جایگزینی بخش‌هایی از جهان تحت مطالعه با یک مدل مشابه‌تر اما با ساختار ساده‌تر است. لذا برای حل بسیاری از مسائل عملی هیدرولوژی ناگزیر به استفاده از مدل‌ها هستیم. اغلب این مدل‌ها بر اساس روابط تعیینی بنا گذاشته شده‌اند و در نتیجه خروجی آنها نیز به صورت یک خروجی منفرد هستند. این در حالی است که به دلیل نادیده گرفتن برخی از فرآیندهای سیستم و ساده‌سازی‌ها خروجی مدل دارای نوعی عدم قطعیت می‌باشد و نیاز است علاوه بر خروجی تعیینی مدل دامنه عدم

Local ) LEC از روش‌های (2014); Dogulu *et al* (2015) (Errors and Clustering) و رگرسیون چندکی (Quantile Regression) جهت برآورد عدم قطعیت مدل‌های هیدرولوژی آب سطحی و در تحقیق (Xu and Valocchi, 2015) از روش‌های جنگل تصادفی و ماشین بردار پشتیبان جهت برآورد عدم قطعیت مدل آب زیرزمینی استفاده شده است.

اگر چه با انتخاب مدل مناسب، استفاده از روش‌های بهینه‌سازی مناسب و همچنین داده‌برداری دقیق‌تر تا حدودی می‌توان عدم قطعیت مربوط به این منابع را کاهش داد، اما همچنان منابع عدم قطعیتی وجود دارد که به‌طور کامل نمی‌توان آنها را نادیده گرفت. در این حالت راهکار عملی برآورد دامنه عدم قطعیت است. بدین منظور در این تحقیق عدم قطعیت و فاصله اطمینان ناشی از عدم قطعیت خروجی مدل HBV که پارامترهای آن با روش تکامل تصادفی جوامع (Shuffled Complex Evolution Metropolis Algorithm) بهینه شده است با استفاده از روش ناپارامتریک LEC کمی می‌شود. روش LEC بر این ایده استوار است که باقیمانده دوره تاریخی (خطا) بین خروجی مدل و داده‌های مشاهداتی بهترین شاخص برای نشان دادن اختلاف بین نتایج مدل و مقادیر واقعی است که اطلاعات ارزشمندی برای ارزیابی عدم قطعیت مدل می‌دهد. در نهایت روش ارائه شده با دو روش ناپارامتریک جنگل تصادفی (Random Forest) و رگرسیون کوانتایل (QR) و روش پارامتریک GLUE در برآورد دامنه عدم قطعیت مقایسه شد.

## مواد و روش‌ها

### منطقه مورد مطالعه

در این مطالعه، ایستگاه هیدرومتری لزوره (کد ۰۵۳-۱۲) با مساحت تحت پوشش ۲۵۶/۸۳ کیلومتر مربع واقع در خروجی حوضه آبخیز چهل‌چای به عنوان ایستگاه مبنا برای شبیه‌سازی بارش-رواناب انتخاب گردید. این حوضه در طول‌های جغرافیایی شرقی ۳۰°، ۲۲°، ۵۵° الی ۳۰°، ۳۷°، ۵۵° و عرض‌های جغرافیایی شمالی ۳۰°، ۵۷°، ۳۶° الی ۱۵°، ۳۷° قرار دارد و از شمال به شهر مینودشت، از غرب به ارتفاعات محمدزمان‌خان و دشت حلقه، از شرق به حوضه آبخیز رودخانه قلی‌تپه و از جنوب به حوضه آبخیز تیل‌آباد محدود می‌گردد (شکل ۱). این منطقه از لحاظ تقسیمات سیاسی در محدوده شهرستان مینودشت در زیرحوضه‌ی بزرگ گرگان‌رود قرار دارد. حداقل ارتفاع حوضه ۱۳۵ متر و حداکثر آن ۲۵۵۰ متر از سطح دریا می‌باشد. ۶۰ درصد از سطح حوضه پوشیده از اراضی جنگلی است و بقیه سطح حوضه را اراضی زراعی در بر می‌گیرد (Rouhani and

روش‌های تقریبی (برای مثال روش‌های مبتنی بر مومنتم اول و دوم) ۳- روش‌های مبتنی بر شبیه‌سازی و نمونه‌برداری ۴- روش‌های بیزی ۵- روش‌های مبتنی بر تجزیه و تحلیل باقیمانده‌ها و ۶- روش‌های مبتنی بر هوش مصنوعی هستند (Solomatine and Shrestha, 2009; Montanari, 2011). در این بین روش‌های تحلیلی برای مدل‌های پیچیده و غیرخطی روابط پیچیده‌ای دارند. همچنین در روش‌های مبتنی بر شبیه‌سازی و بیزی نیاز است تا تابع توزیع آماری مناسب انتخاب گردد. در این حالت صحت نتایج حاصل از برآورد عدم قطعیت بستگی به صحت فرضیات در نظر گرفته شده برای توابع توزیع پیشین دارد. علاوه بر این در بسیاری از روش‌ها فقط عدم قطعیت ناشی از یک منبع بررسی می‌شود. به عنوان مثال در روش مونت کارلو اغلب عدم قطعیت ناشی از ورودی‌ها (Khu and Werner, 2003) و در روش بیزی عدم قطعیت ناشی از پارامترها (Houska *et al*, 2014) بررسی می‌گردد. این در حالی است که معمولاً مدیران و تصمیم‌گیران منابع آب نیاز به عدم قطعیت کلی خروجی مدل دارند. در این حالت برآورد عدم قطعیت کلی مدل با استفاده از بررسی هر یک از منابع عدم قطعیت از نظر روابط و محاسبات امری مشکل است و در برخی موارد محاسبه این عدم قطعیت ممکن نمی‌باشد. در حالی در رویکردهایی که عدم قطعیت با تجزیه و تحلیل باقیمانده‌ها برآورد می‌شود، کمتر با این مشکل مواجه هستند. در این حالت فرض می‌شود که عدم قطعیت پارامترها، ورودی و ساختار مدل به نوعی در مقادیر باقیمانده‌ها خلاصه شده است (Xu and Valocchi, 2015). در این ارتباط (Evin *et al* (2014) مشاهده کردند که نتایج روش‌های مبتنی بر تجزیه و تحلیل باقیمانده‌ها در مقایسه با نتایج روش‌هایی که اثر تمام منابع عدم قطعیت را در نظر می‌گیرند، قابل قبول‌تر است. در روش‌های مبتنی بر تجزیه و تحلیل باقیمانده‌ها نیازی به برخی فرضیات (مثل نرمال بودن باقیمانده‌ها، عدم خود همبستگی بین مقادیر باقیمانده‌ها) در مورد باقیمانده‌ها نمی‌باشد. در واقع این روش‌ها بر اساس تجزیه و تحلیل مقادیر باقیمانده‌ها قادرند روابط پیچیده بین خروجی (به عنوان مثال مقادیر باقیمانده یا صدک‌های آن) و ورودی‌ها را مدل‌سازی کنند (Xu and Valocchi, 2015). در این روش‌ها فرض می‌شود خطایی (مقادیر باقی‌مانده‌ها) که در زمان حال اتفاق افتاده در زمان‌های آینده با شرایط یکسان نیز اتفاق خواهد افتاد در این حالت از خطای اتفاق افتاده در زمان حال برای تعیین عدم قطعیت خروجی مدل در آینده استفاده می‌گردد (López López *et al*, 2014). در تحقیقات (Solomatine and Shrestha (2009); Weerts *et al* (2011); López López *et al*

(که این پارامتر با توجه به داده‌های مشاهداتی بهینه می‌شود) باشد تبخیر واقعی به صورت تابعی خطی از تبخیر پتانسیل در نظر گرفته می‌شود و چنانچه ارتفاع آب بیشتر از حد LP باشد تبخیر واقعی برابر با تبخیر پتانسیل در نظر گرفته می‌شود. آن قسمت از بارش که در سطح خاک جریان ندارد به ذخیره آب خاک اضافه می‌شود. در بخش رواناب، خروجی اصلی مدل یعنی جریان خروجی حوضه که خود شامل سه بخش جریان سطحی، زیرسطحی و جریان پایه است، محاسبه می‌شود. در این قسمت برای تعیین دبی خروجی از دو مخزن استفاده می‌شود. اولین مخزن نزدیک سطح خاک بوده و برای محاسبه رواناب سطحی و زیر قشری استفاده می‌شود و دومین مخزن در عمق بیشتر قرار گرفته و برای شبیه‌سازی جریان پایه استفاده می‌شود. آن قسمت از آب باران و ذوب برف که بیشتر از ظرفیت نگهداری آب خاک است وارد مخزن اول می‌شود. هر دو مخزن در ارتباط با هم بوده و با توجه به ذخیره آب هر مخزن، آستانه ارتفاع در مخزن و ضرایب خشکیدگی بخش مختلف رواناب محاسبه می‌شود (AghaKouchak and Habib, 2010). ورودی‌های مدل HBV شامل داده‌های بارش، دما و تبخیر و تعرق پتانسیل هستند و در مجموع مدل دارای ۱۳ پارامتر است که نیاز است تا با روش مناسب این پارامترها بهینه شوند.

#### روش‌های برآورد عدم قطعیت

فرض می‌شود با بردار ورودی‌های  $x$  و پارامترهای  $\theta$ ، مدل تعیینی  $M$  (که در اینجا مدل HBV است) دارای خروجی  $\hat{y}$  است. منابع خطای مختلف (ورودی‌ها، ساختار و پارامترهای مدل) بر خروجی مدل تأثیر می‌گذارند در این حالت خروجی سیستم می‌تواند به صورت رابطه (۱) بیان گردد.

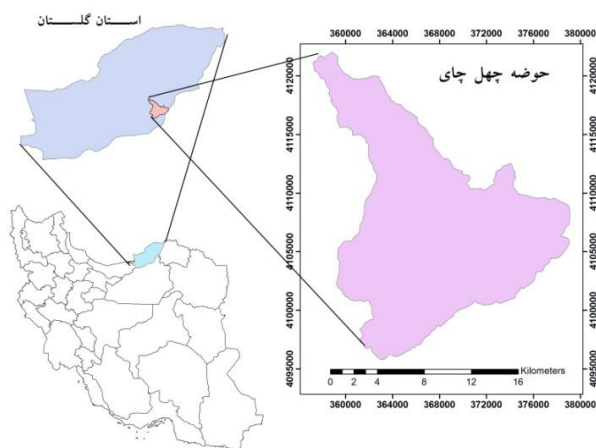
$$y_t = \hat{y} + e = M(x, \theta) + e \quad (\text{رابطه ۱})$$

که در اینجا  $e$  مقدار کل خطای مدل و  $t$  واحد زمان است. در روش‌های مبتنی بر داده فرض می‌شود عدم قطعیت منابع مختلف در مقادیر باقیمانده‌ها خلاصه شده است؛ بنابراین با بررسی باقیمانده‌ها و تجزیه و تحلیل آنها می‌توان دامنه عدم قطعیت خروجی مدل را بدست آورد.

#### روش LEC

این روش اولین بار توسط Shrestha and Solomatine (2006) معرفی شد. در این روش یک مدل رگرسیونی غیرخطی برای تخمین چندک‌های توزیع باقیمانده‌ها ساخته می‌شود و فرض می‌شود که عدم قطعیت خروجی مدل به خصوصیات سیستم و مقادیر باقیمانده‌ها بستگی دارد (Dogulu et al, 2015). برای اجرای روش LEC مراحل کار به شرح زیر است:

(Farahi Moghadam, 2014). در این تحقیق داده‌های پایه مورد استفاده شامل داده‌های مشاهداتی بارندگی، دبی رودخانه و تبخیر تعرق روزانه می‌باشد. از سال آبی ۱۳۶۷-۶۸ تا سال آبی ۱۳۷۱-۷۲ برای دوره واسنجی مدل و از سال آبی ۱۳۷۲-۷۳ تا سال ۱۳۷۳-۷۴ برای دوره اعتبارسنجی انتخاب گردید. رودخانه مورد نظر دائمی می‌باشد که دبی متوسط سالیانه آن در دوره‌های واسنجی و اعتبارسنجی به ترتیب ۲/۷۱ و ۲/۶۲ متر مکعب بر ثانیه است.



شکل ۱. موقعیت حوضه چهل‌چای

#### روش تحقیق

مدل هیدرولوژیکی HBV (Hydrologiska Byrans Vattenavdelning) مدل HBV یک مدل بارش-رواناب یکپارچه مفهومی جهت شبیه‌سازی فرآیندهای هیدرولوژی حوضه آبخیز است که توسط موسسه مطالعات هیدرولوژی و هواشناسی سوئد (Swedish Meteorological and Hydrological Institute) برای پیش‌بینی مقدار رواناب سیستم‌های تولید انرژی در سال ۱۹۷۰ ارائه شده است. اطلاعات ورودی این مدل علاوه بر کم بودن، در دسترس نیز هستند. همچنین در شرایط آب و هوایی متفاوت، این مدل قابلیت استفاده دارد (Siebert and Vis, 2012). این مدل شامل سه بخش اصلی: ۱- بخش ذوب و تجمع برف ۲- بخش رطوبت خاک و بارش مؤثر ۳- بخش رواناب، می‌باشد. در بخش ذوب و تجمع برف در هر گام زمانی با توجه به دمای مشاهداتی و دمای آستانه  $T$ ، بارش دریافتی یا به صورت باران یا به صورت برف در نظر گرفته می‌شود. تا زمانی که دما کمتر از دمای آستانه باشد برف تجمع کرده و هیچ‌گونه روانابی تشکیل نمی‌شود. آب حاصل از ذوب برف و باران وارد بخش خاک شده و با توجه به رطوبت خاک و همچنین حداکثر قابلیت ذخیره رطوبت توسط خاک، مقدار روانابی که در سطح خاک جریان پیدا می‌کند محاسبه می‌شود. در این بخش چنانچه ذخیره آب خاک از ۰ تا حد LP

بهینه‌سازی پارامترهای مدل هیدرولوژیکی. در این قسمت پارامترهای مدل HBV توسط روش بهینه‌سازی SCE\_MA، بهینه می‌شوند. در این حالت با انتخاب پارامترهای بهینه قسمتی از عدم قطعیت مدل ناشی از پارامترها از بین می‌رود. اجرای مدل هیدرولوژیکی بهینه شده برای دوره واسنجی و صحت سنجی و بدست آوردن مقادیر باقیمانده‌ها.

تعیین شرایط هیدرولوژیکی مشابه و مهمترین متغیرهایی که بر روی باقیمانده‌ها مؤثر هستند. در روش LEC فرض می‌شود در شرایط هیدرولوژیکی مشابه مقادیر باقیمانده‌های مدل از لحاظ بزرگی و توزیع شبیه یکدیگر هستند. بنابراین در ابتدا شرایط مشابه هیدرولوژیکی برای دوره‌های زمانی مختلف تعیین می‌شود. در مدل‌های هیدرولوژیکی شرایط هیدرولوژیکی به وسیله مقادیر متغیرهایی مانند بارندگی، دبی و غیره تعیین می‌گردد. بنابراین ابتدا بهترین ورودی‌ها که تبیین‌کننده شرایط هیدرولوژیکی مشابه با مقادیر باقیمانده‌های شبیه به هم هستند، تعیین می‌شوند. بدین منظور در این تحقیق مقادیر دبی روزانه مشاهداتی، دبی روزانه شبیه‌سازی شده، باقیمانده‌ها در گام زمانی قبل، بارندگی روزانه و تبخیر و تعرق روزانه از یک تا پنج گام زمانی قبل به عنوان متغیرهای محتمل برای تعیین شرایط هیدرولوژیکی مشابه در نظر گرفته شدند. سپس با استفاده از مقادیر همبستگی بین ورودی‌ها و باقیمانده‌ها در هر گام زمانی و همچنین تابع IMPORTANCE جنگل تصادفی بهترین متغیرهای ورودی تعیین شدند.

تعیین تعداد خوشه‌های فازی و خوشه‌بندی ورودی‌ها و باقیمانده‌های متناظر آنها. پس از تعیین ورودی‌های مناسب با استفاده از روش خوشه‌بندی فازی، مقادیر عضویت هر یک از بردارهای ورودی و مقادیر باقیمانده‌های متناظر آنها برای هر گام زمانی در خوشه‌های مختلف فازی محاسبه می‌شود. در روش خوشه‌بندی فازی مسئله مهم تعیین تعداد خوشه‌ها است که در این تحقیق از معیارهای  $V_{pc}$  (Partition Coefficient)،  $V_{pe}$  (Partition Entropy) (Bezdek 1974 a,b)،  $V_{XB}$  (Xie and Beni, 1991) و  $V_{fs}$  (Sugeno and Fukuyama and Sugeno, 1989) برای تعیین تعداد خوشه بهینه استفاده شد. علاوه بر معیارهای فوق با توجه به آن که در این تحقیق هدف برآورد دامنه عدم قطعیت است از معیارهای ارزیابی مربوط به دامنه عدم قطعیت شامل معیارهای MPI (Mean Prediction Interval) و PICP (Prediction Interval Coverage Probability) نیز برای تعیین تعداد خوشه بهینه استفاده گردید (Shrestha and Solomatine, 2006; Solomatine and Shrestha, 2009). پس از تعیین تعداد خوشه بهینه، اقدام به

خوشه‌بندی بردارهای ورودی و باقیمانده‌های متناظر آنها گردید. تهیه تابع توزیع تجمعی تجربی و برآورد چندک‌های باقیمانده‌های در هر خوشه. در این گام با تجزیه و تحلیل مقادیر باقیمانده‌ها در هر خوشه تابع توزیع تجمعی تجربی برای مقادیر باقیمانده‌ها در هر خوشه برآورد و چندک‌های مختلف محاسبه می‌شوند. در روش خوشه‌بندی فازی هر داده با درجه عضویت‌های مختلف به چند خوشه فازی تعلق دارد؛ بنابراین نمی‌توان چندک‌ها را فقط با در نظر گرفتن مقادیر باقیمانده‌های هر خوشه محاسبه کرد؛ بلکه باید مقادیر عضویت باقیمانده‌ها در هر خوشه فازی را نیز در محاسبه مقادیر چندک‌ها در نظر گرفت. بدین منظور در ابتدا مقادیر باقیمانده‌ها بصورت صعودی مرتب می‌شوند. سپس با استفاده از رابطه‌های (۲) و (۳) مقادیر چندک (در این تحقیق چندک ۰/۰۵) مربوط به حد پایین برآورد می‌گردد (Shrestha and Solomatine, 2006).

$$PIC_i^L = e_j \quad (\text{رابطه ۲})$$

$$j: \sum_{k=1}^j \mu_{i,k} < a/2 \sum_{j=1}^n \mu_{i,j} \quad (\text{رابطه ۳})$$

در این رابطه  $j$  حداکثر رتبه باقیمانده‌ها است که نامعادل بالا را برآورد کند،  $e_j$  مقدار باقیمانده مربوط به نمونه  $j$ ،  $\mu_{i,j}$  مقدار عضویت  $i$  از نمونه در خوشه  $i$  است.  $a$  سطح معنی‌داری (در این تحقیق سطوح معنی‌داری ۰،۰۲۵ و ۰،۹۷۵ مورد بررسی قرار گرفتند) و  $n$  تعداد کل نمونه‌هایی است که در خوشه  $i$  دارای عضویت هستند. برای حد بالا نیز مشابه رابطه‌های (۲) و (۳) استفاده می‌گردد.

محاسبه چندک‌های باقیمانده‌ها در سطوح معنی‌داری مورد نظر برای داده‌های دوره واسنجی. چندک‌های هر خوشه با استفاده از رابطه‌های (۲) و (۳) محاسبه می‌شوند. ولی چنانچه هدف محاسبه دامنه عدم قطعیت برای هر بردار ورودی باشد نیاز است تا مقادیر چندک‌ها برای هر بردار ورودی داده‌های دوره واسنجی محاسبه گردد. در صورتی که داده‌ها با استفاده از خوشه‌بندی فازی کلاسه‌بندی شده باشند برای محاسبه چندک‌ها برای هر بردار ورودی از رویکرد Fuzzy Committe استفاده می‌شود (Solomatine and Siek, 2006) (رابطه ۴).

$$e_t^p = \sum_{i=1}^c \mu_{i,t}^{2/m} ec_i^p \quad (\text{رابطه ۴})$$

$e_t^p$  مقدار چندک  $p$  باقیمانده‌ها برای بردار ورودی زمان  $t$ ،  $ec_i^p$  مقدار چندک  $p$  خوشه  $i$  مقدار ضریب صاف‌کاری توانی،  $\mu_{i,t}$  مقدار عضویت برداری ورودی زمان  $t$  در خوشه  $i$  و  $c$  تعداد کل خوشه‌ها هستند. به عبارت دیگر در رابطه (۴) برای بدست آورد مقدار چندک هر بردار ورودی، متناسب با عضویت بردار ورودی در کلاس‌های مختلف از مقادیر چندک‌ها متوسط‌گیری می‌شود.

چندکها یک برآورد کامل از تابع توزیع متغیر وابسته (که به نوعی نماینگر مقدار عدم قطعیت آن پارامتر است) بدست می آید (Weerts et al, 2011). در تحقیق (Lopez Lopez et al (2014) چهار مدل رگرسیون چندک شامل مدل با داده های اولیه، مدل با تبدیل چندکی نرمال (Normal quantile transformation)، مدل در حالتی که مسئله تقاطع چندکها (Non-crossing) (quantile regression) مورد توجه قرار گرفته است و مدل رگرسیون چندکی خطی قطعه قطعه (Piecewise linear) (quantile regression)، برای برآورد عدم قطعیت در رودخانه سورن (Severn) مقایسه و ارزیابی گردید. آنها بیان کردند که عملکرد هر چهار روش نسب به یکدیگر تفاوتی چندانی نداشته اند. بنابراین در این تحقیق ساده ترین مدل که ورودی آن مقادیر دبی برآوردی بدون نرمال سازی است، استفاده گردید. در روش رگرسیون چندکی برای هر چندک  $\tau$  یک رابطه خطی بین مقادیر مشاهداتی ( $y$ ) و شبیه سازی شده ( $\hat{y}$ ) بصورت رابطه ( $\gamma$ ) برقرار است.

$$y = a_{\tau}\hat{y} + b_{\tau} \quad (\text{رابطه } \gamma)$$

در رابطه ( $\gamma$ )،  $a_{\tau}$  و  $b_{\tau}$  پارامترهای مدل رگرسیون چندکی هستند که با استفاده از حداقل کردن جمع مربعات خطا (رابطه ۸) برآورد می شوند.

$$\min \sum_{j=1}^J p_{\tau}(y_j - (a_{\tau}\hat{y}_j + b_{\tau})) \quad (\text{رابطه } ۸)$$

در اینجا  $y_j$  و  $\hat{y}_j$  مقادیر زامین نمونه از کل نمونه ها و  $p_{\tau}$  تابع رگرسیون چندکی برای چندک  $\tau$  هستند.

$$p_{\tau}(\epsilon_j) = \begin{cases} (\tau - 1) \cdot \epsilon_j, & \epsilon_j \leq 0 \\ \tau \cdot \epsilon_j, & \epsilon_j \geq 0 \end{cases} \quad (\text{رابطه } ۹)$$

رابطه (۹) برای باقیمانده ( $\epsilon_j$ ) که به عنوان تفاوت بین مقادیر مشاهداتی و برآوردی برای چندک  $\tau$  است، تعریف می شود. در این تحقیق از پکیج quantreg نرم افزار R برای اجرای رگرسیون چندک استفاده شد.

#### جنگل تصادفی (RF)

روش درخت تصمیم یکی از روش های ناپارامتریک مناسب برای کلاسه بندی و رگرسیون است. از مزایای این روش ساده بودن فهم آن و عدم نیاز به فرض توزیع آماری برای داده ها است. از معایب این روش نیز عدم پایایی آن می باشد؛ بطوری که کوچکترین تغییر در داده های ورودی منجر به تغییر گسترده در ساختار درخت و خروجی مدل می شود. جهت رفع این مشکل (Breiman (2001)، جنگل تصادفی که بر اساس ایجاد تعداد زیادی درخت تصمیم ارائه داد. در این روش هر درخت تصمیم بر روی نمونه بوت استرپ (Bootstrap) که با نمونه برداری همراه با جایگذاری داده اولیه بدست آمده است، توسعه داده می شود.

محاسبه چندک های باقیمانده ها در سطوح معنی داری مورد نظر برای داده های دوره صحت سنجی. در گام قبل مقادیر چندکها برای هر بردار ورودی دوره واسنجی محاسبه گردید. در این گام به منظور محاسبه مقادیر چندکها برای بردارهای ورودی مربوط به دوره صحت سنجی در ابتدا با استفاده از مدل هوش مصنوعی (رابطه ۵) یک رابطه بین چندکها و ورودیها در دوره واسنجی برقرار می شود، سپس با اجرای مدل برای بردارهای ورودی دوره صحت سنجی مقادیر چندکها برای داده های دوره صحت سنجی محاسبه می شوند.

$$e^p = U^p(X_{it}, \theta^p) \quad (\text{رابطه } ۵)$$

در اینجا  $\theta^p$  بردار پارامترهای مدل  $U^p$  برای  $p$  امین چندک است.

محاسبه دامنه عدم قطعیت برای مقادیر دبی در دوره واسنجی و صحت سنجی. پس از محاسبه مقادیر چندک های باقیمانده ها برای بردارهای ورودی دوره واسنجی و صحت سنجی، در نهایت دامنه عدم قطعیت برای خروجی مدل به صورت رابطه (۶) محاسبه می شود.

$$y^p = \hat{y} + e^p \quad (\text{رابطه } ۶)$$

در اینجا  $y^p$  چندک  $p$  خروجی مدل (دبی شبیه سازی شده) است. به منظور برآورد برای مثال فاصله اطمینان ۹۰ درصد نیاز است تا مدل های  $U^{95}$  (صدک ۹۵ درصد) و  $U^5$  (صدک ۵ درصدی) ساخته شوند.

#### رگرسیون چندک (QR)

رگرسیون چندک یک روش آماری با قابلیت محاسبه و رسم منحنی های رگرسیونی متفاوت و منطبق با نقاط چندکی مختلف است، این روش ضمن ارائه تصویری کامل و جامع از داده ها، امکان سنجش ارتباط بین متغیرهای مستقل با چندک های مورد نظر متغیر وابسته را بدون نیاز به نرمال بودن داده ها و حتی در حضور نقاط دور افتاده فراهم می کند. در روش رگرسیون چندک سعی می شود تا تابع توزیع احتمال متغیر وابسته با استفاده از یک یا تعداد بیشتری متغیر مستقل تشریح شود (Koenker and Bassett, 1978; Koenker, 2005). از آنجایی که این رگرسیون بر اساس چندکها است، داده های پرت نمی توانند آن را تحت تأثیر قرار دهند. همچنین در این روش نیازی به فرض پیروی داده ها از توزیع مشخص مثل توزیع نرمال، بتا و غیره نمی باشد. بنابراین به نوعی یک روش ناپارامتریک است. در مقایسه با روش رگرسیون معمولی در رگرسیون چندک مقدار میانگین متغیر وابسته برآورد نمی گردد بلکه مقادیر چندکها برآورد می گردد. بنابراین با برآورد

برآوردی با سطح معنی داری  $1 - \alpha$  به عنوان مثال ۹۰ درصدی است.

$$PICP = \frac{1}{n} \sum_{t=1}^n C \quad (\text{رابطه ۱۰})$$

$$\left\{ \begin{array}{l} 1, PL_t^{lower} \leq y_t \leq PL_t^{upper} \\ 0, \text{ سایر حالت ها} \end{array} \right\} \quad (\text{رابطه ۱۲})$$

در حالت ایده آل مقدار PICP برابر یا نزدیک مقدار فاصله اطمینان است. معیار MPI مقدار متوسط عرض دامنه عدم قطعیت را محاسبه می کند. به عنوان مثال فاصله بین حد بالا و پایین دامنه عدم قطعیت برابر است با:

$$MPI = \frac{1}{n} \sum_{t=1}^n (PL_t^{upper} - PL_t^{lower}) \quad (\text{رابطه ۱۳})$$

مقدار  $MPI=0$  به مفهوم عدم وجود عدم قطعیت است. معیار ARIL مشابه MPI است و در هر گام زمانی مقدار متوسط عرض دامنه عدم قطعیت را نسبت به مقدار دبی مشاهداتی در نظر می گیرد.

$$ARIL = \frac{1}{n} \sum_{t=1}^n \frac{(PL_t^{upper} - PL_t^{lower})}{y_t} \quad (\text{رابطه ۱۴})$$

در حقیقت در رابطه (۱۲) به این دلیل در مخرج مقدار مشاهداتی استفاده شده است که انتظار می رود با افزایش مقدار دبی مقدار عدم قطعیت برآورد آن نیز افزایش یابد.

### نتایج و بحث

پس از بهینه سازی پارامترهای مدل یکپارچه هیدرولوژیکی HBV با استفاده از روش SCE\_MA (از آنجا که تاکید مقاله بر روی بررسی عدم قطعیت بود نتایج آن ارائه نشده است) خروجی مدل برای دوره واسنجی و صحت سنجی بدست آمد که نتایج آن به صورت شکل (۲) است. با توجه به شکل (۲) مشاهده می گردد که مدل در برآورد مقادیر داده های کمینه عملکرد بهتری نسبت به برآورد داده های بیشینه داشته است. در تحقیق Dogulu et al (2015) در مدل سازی بارش- رواناب با استفاده از مدل HBV عملکرد مدل در برآورد مقادیر دبی کمینه بهتر از مقادیر دبی بیشینه بود. در شکل (۳) مقادیر باقیمانده ها برای دوره واسنجی و صحت سنجی ارائه شده است. شکل (۳) نشان می دهد با افزایش مقدار دبی مقادیر باقیمانده ها نیز بزرگتر شده اند و به نوعی بیانگر ناهمگنی مقادیر باقیمانده ها ناهمگن می باشند. همچنین بررسی نرمال بودن مقادیر باقیمانده ها نیز بیانگر نرمال نبودن باقیمانده ها است. این در حالی است که در اغلب روش های پارامتری برآورد عدم قطعیت نیاز است تا داده ها نرمال و همگن باشند. در تحقیقات Solomatine and Shrestha (2009); Dogulu et al (2015) نیز مشاهده کردند مقادیر باقیمانده ها همگن نبوده و از توزیع نرمال پیروی نمی کنند.

در هر بار نمونه گیری با جایگزینی از اطلاعات، برخی از اطلاعات هرگز نمونه گیری نمی شوند و برخی دیگر شاید چند بار نمونه گیری شوند. به عبارت دیگر برخی داده های ورودی برای برخی درخت ها در اصطلاح نمونه خارج از کیسه (Out of bag) خواهند بود یعنی در ایجاد برخی درخت ها مشارکت نخواهند داشت. پس از آموزش مدل برای پیش بینی داده تست  $X^*$  تمام درخت های تصمیم با در نظر گرفتن این ورودی ها اجرا شده و در نهایت با میانگین گیری نتایج تمام درخت های تصمیم ساخته شده خروجی متوسط مدل بدست می آید. به دلیل اینکه بیش برآزشی در درخت تصمیم اتفاق نمی افتد بنابراین نیازی به هرس کردن هر درخت تصمیم نمی باشد. همچنین برای هر تقسیم از بین متغیرهای موجود، متغیر تقسیم شوند به صورت تصادفی انتخاب می گردد. در این روش دو پارامتر مهم تعداد متغیرهای انتخابی برای تقسیم ((ارامتر M) و حداقل تعداد برگ ها در هر شاخه هستند که در این تحقیق مقدار پارامتر M برابر با یک سوم تعداد متغیرهای ورودی مشاهداتی و حداقل تعداد برگ ها در هر شاخه به صورت سعی و خطا تعیین گردید. برای اجرای روش جنگل تصادفی از پکیج randomforest نرم افزار R استفاده گردید. در این تحقیق برای برآورد دامنه عدم قطعیت، مقادیر باقیمانده ها (اختلاف بین مقادیر دبی مشاهداتی و دبی شبیه سازی شده با مدل HBV) به عنوان خروجی مدل جنگل تصادفی و متغیرهایی که در روش LEC به عنوان بهترین ورودی ها انتخاب شده بودند به عنوان ورودی مدل جنگل تصادفی در نظر گرفته شدند. با در نظر گرفتن تعداد شبیه سازی ها برابر با ۲۵۰۰، تعداد ۲۵۰۰ درخت تصمیم با استفاده از داده های مرحله واسنجی ساخته شد. سپس مدل برای ورودی های مرحله صحت سنجی اجرا گردید و برای هر بردار ورودی مرحله صحت سنجی، خروجی مربوط به ۲۵۰۰ درخت تصمیم بدست آمد. در گام بعد با استفاده از ۲۵۰۰ خروجی درخت های تصمیم، تابع توزیع تجمعی تجربی محاسبه و قادر چندک های ۰/۵ و ۰/۹۵ باقیمانده های محاسبه گردید. در نهایت با در نظر گرفتن این چندک های محاسباتی و همچنین مقادیر دبی شبیه سازی شده با مدل HBV و رابطه (۶)، دامنه عدم قطعیت برای مقادیر دبی شبیه سازی شده بدست آمد.

### معیارهای ارزیابی

به منظور مقایسه ارزیابی و مقایسه روش های مختلف برآورد عدم قطعیت از معیارهای PICP, MPI و Average Relative ARIL (Interval Length) استفاده شد. PICP نشان دهنده احتمال قرارگیری مقدار مشاهداتی ( $y_t$ ) در داخل دامنه عدم قطعیت

رگرسیون بردار پشتیبان با تابع کرنل گوسی برای صدک های ۵ و ۹۵ درصد اجرا گردید و با استفاده از روش جستجوی شبکه ای بهترین مجموعه پارامترها تعیین گردید. در نهایت با استفاده از مقادیر پارامترهای بهینه، مدل برای دوره صحت سنجی اجرا گردید و صدک های ۵ و ۹۵ درصدی باقیمانده ها محاسبه شدند و در نهایت با استفاده از رابطه (۶) فاصله اطمینان ۹۰ درصدی برآورد گردید که نتایج آن در شکل (aV) ارائه شده است. با توجه شکل (aV) مشاهده می شود بیشترین داده های خارج از دامنه عدم قطعیت مربوط به دبی های سیلابی است. برای ماه های کم آبی در اغلب موارد حد پایین صفر بدست آمده است و در ماه های پر آبی داده مشاهداتی بیشتر نزدیک به حد بالا هستند. در تحقیقات (Solomatine and Shrestha (2009); Dogulu et al (2015) جهت برآورد چندک های مختلف از یک مدل هوش مصنوعی استفاده کردند. در این تحقیق علاوه بر استفاده از روش هوش مصنوعی از رویکردی که مبتنی بر خوشه بندی فازی است استفاده شد. در این روش با استفاده پارامترهای خوشه بندی فازی که برای دوره واسنجی محاسبه گردیده، مقادیر عضویت بردارهای ورودی مربوط به دوره صحت سنجی در کلاس های مختلف محاسبه گردید. سپس با استفاده از چندک های محاسبه شده در جدول (۱) و رابطه (۴) مقادیر چندک های باقیمانده ها برای هر بردار ورودی و در نهایت با استفاده از رابطه (۶) و مقادیر دبی شبیه سازی شده توسط مدل HBV، مقادیر چندک ها و دامنه عدم قطعیت دبی خروجی محاسبه گردید که نتایج آن در شکل (bV) ارائه شده است. با توجه به این شکل ملاحظه می شود عملکرد این روش شبیه عملکرد روش در حالتی که ماشین بردار رگرسیون استفاده شده است می باشد و اختلاف بین شکل ها بسیار جزئی هستند. بنابراین با توجه به محاسبات کمتر و عملکرد مناسب روش LEC، در حالتی که آموزش ندیده نسبت به حالتی که با مدل هوش مصنوعی آموزش داده شده باشد، استفاده از این روش پیشنهاد می گردد.

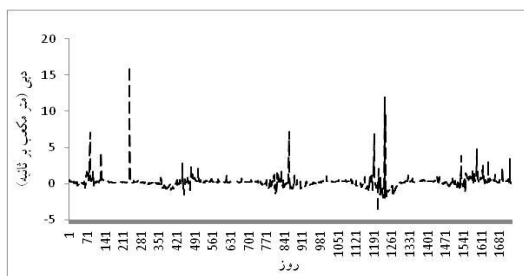
جدول ۱. مقادیر چندک ها در خوشه های مختلف

صدک ۹۵٪	صدک ۵٪	
۳/۱۲	-۳/۰۱	خوشه ۱
۶/۸۶	-۴/۵۶	خوشه ۲
۳/۶۴	-۲/۵۶	خوشه ۳
۱/۵۶	-۱/۲۷	خوشه ۴

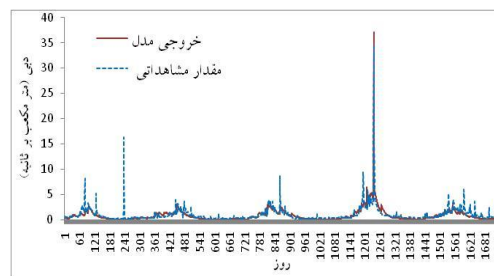
در گام بعد با استفاده از خوشه بندی فازی، باقیمانده ها از نظر شرایط مختلف هیدرومتئولوژیکی خوشه بندی شدند. برای این منظور ابتدا در روش LEC مهمترین پارامترهای ورودی با استفاده از ضریب همبستگی و تابع Importance جنگل تصادفی تعیین شدند که در این قسمت متغیرهای بارش و دبی برآوردی همان گام، دبی مشاهداتی یک گام زمانی قبل، باقیمانده ها در یک و سه گام زمانی قبل به عنوان مهمترین متغیرها انتخاب شدند. روش LEC یک روش موضعی است در این حالت چنانچه دامنه عدم قطعیت برآورد گردد دارای نوسانات زیاد است. جهت رفع این مشکل و افزایش خودهمبستگی بین مقادیر حدهای بالا و پایین دامنه عدم قطعیت از مقادیر ورودی های دبی برآوردی و مشاهداتی و بارندگی تا چهار گام زمانی متوسط گیری شد و به عنوان ورودی های مدل استفاده گردید.

پس از شناسایی متغیرهای مهم در روش خوشه بندی فازی با استفاده از شاخص های ارزیابی اقدام به تعیین تعداد خوشه بهینه گردید. نتایج مربوط به شاخص های مختلف ارزیابی در شکل (۴) ارائه شده است. با توجه به شکل (۴) مشاهده می شود با افزایش تعداد خوشه مقدار تمام شاخص ها به غیر از شاخص VFS افزایش می یابد که بیانگر عدم توانایی این شاخص ها در تعیین تعداد خوشه بهینه می باشد. بنابراین در ادامه از شاخص های MPI و PICP برای تعیین تعداد خوشه بهینه استفاده گردید که نتایج این دو روش در شکل (۵) ارائه شده است؛ این شکل نشان می دهد که با افزایش تعداد خوشه، شاخص های MPI و PICP به ترتیب کاهش و افزایش پیدا می کنند. در تعداد خوشه برابر با چهار یک حالت تعادلی بین این دو شاخص برقرار است؛ بنابراین چهار خوشه به عنوان تعداد خوشه بهینه انتخاب گردید. در تحقیقات (Shrestha and Dogulu et al (2015); Solomatine (2009) تعداد ۶ خوشه به عنوان تعداد خوشه بهینه تشخیص داده شدند.

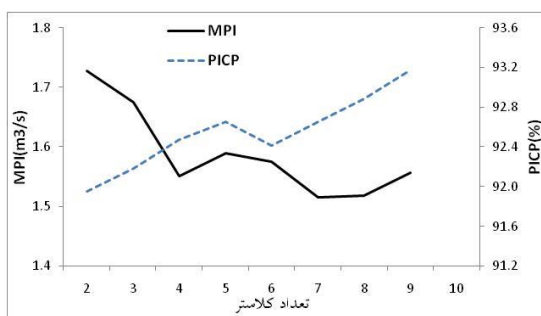
شکل (۶) قسمت های مختلف هیدروگراف را که در خوشه های مختلف قرار گرفته اند نشان می دهد. مطابق شکل (۶) مشاهده می شود که در خوشه های ۱ و ۳ به ترتیب دبی های مربوط به ماه های کم و پر آبی قرار می گیرند، در خوشه ۴ دبی های سیلابی و در خوشه ۲ مقادیر دبی در دامنه های مختلف که در سایر خوشه ها نمی باشند قرار گرفته اند. پس از برازش توزیع تجربی به مقادیر باقیمانده های هر خوشه مقادیر صدک های ۵ و ۹۵ درصدی هر خوشه محاسبه گردید که در جدول (۱) آورده شده است. با در نظر گرفتن ورودی ها، مدل



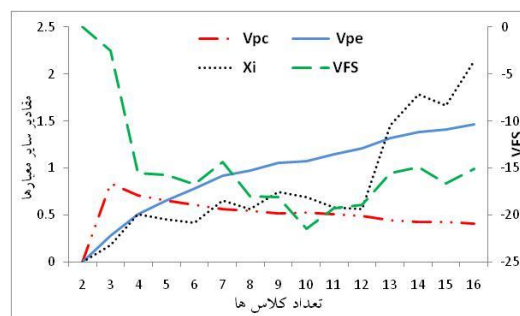
شکل ۳. مقادیر باقیمانده‌ها برای گام‌های زمانی مختلف HBV



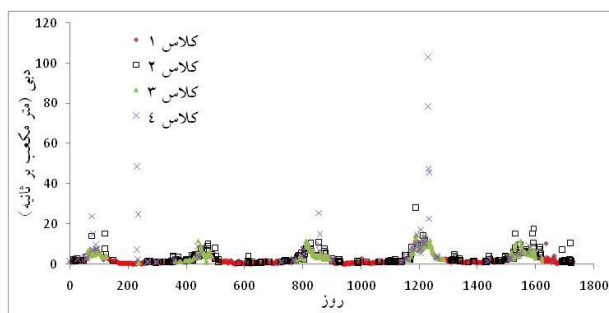
شکل ۴. مقدار مشاهداتی و برآورد شده دبی با استفاده از مدل



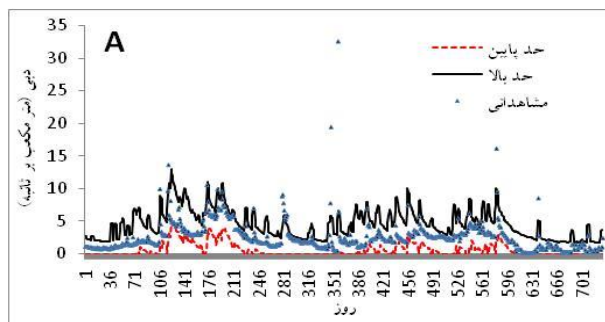
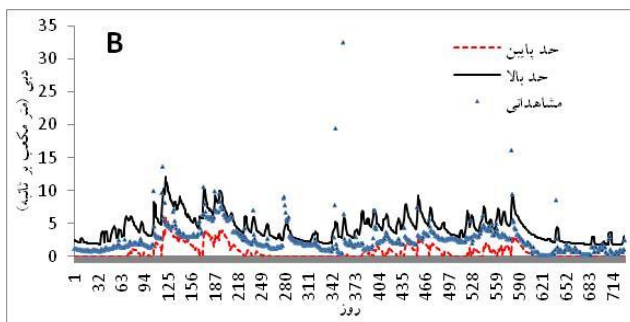
شکل ۵. مقادیر شاخص‌های MPI و PICP برای تعداد کلاس‌های مختلف



شکل ۶. مقادیر معیارهای ارزیابی برای تعداد کلاس‌های مختلف



شکل ۷. مقادیر دبی‌ها در کلاس‌های مختلف



شکل ۸. دامنه عدم قطعیت برآوردی a: روش LEC در حالتی که آموزش دیده باشد b: روش LEC در حالتی که آموزش داده نشده است

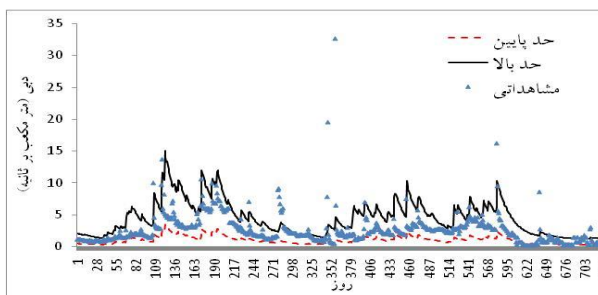
در روش رگرسیون چندک فقط از مقادیر دبی شبیه‌سازی شده به عنوان ورودی مدل استفاده و مقادیر صدک‌های ۵ و ۹۵ درصد برای خروجی محاسبه گردید که در شکل (۸) دامنه عدم قطعیت به همراه مقادیر دبی مشاهداتی ارائه شده است. با توجه به شکل (۸) مشاهده می‌شود که بر خلاف روش LEC که در

بسیاری موارد مقدار حد پایین را برابر با صفر برآورد کرده بود در این روش مقادیر حد پایین بزرگتر از صفر بدست آمده است. در صورتی که مانند روش LEC در دبی‌های بالا مقدار مشاهداتی نزدیک به حد بالا و در دبی‌های پایین داده‌های مشاهداتی نزدیک حد پایین بوده‌اند. همچنین در این روش نیز مانند روش

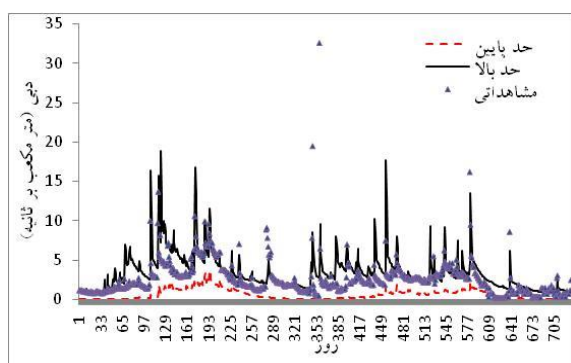


منحنی‌های حد بالا و حد پایین، کاسته شده و فاصله بین حد پایین و بالا نیز کاهش پیدا می‌کند، در این حالت تعداد زیادی از مشاهدات خارج از دامنه عدم قطعیت قرار گرفتند. در تعداد ۵۰۰ برگ بین مقدار فاصله حد بالا و پایین و نوسانات منحنی‌های حد بالا و حد پایین یک حالت تعادل برقرار گردید و این تعداد به عنوان مقدار بهینه انتخاب گردید. در نهایت مدل با تعداد ۵۰۰ برگ در هر شاخه برای دوره صحت‌سنجی اجرا گردید که نتایج آن در شکل (۹) ارائه شده است. با توجه به این شکل ملاحظه می‌شود در این روش نسبت به دو روش دیگر فاصله بین حد پایین و حد بالا کمتر می‌باشد. در این روش نوسانات مربوط به حد پایین کم بوده، در حالی که برای حد بالا نوسانات منحنی عدم قطعیت بدست آمده بیشتر شده است و برای برخی مقادیر دبی حد بالای خیلی بزرگی بدست آمده است.

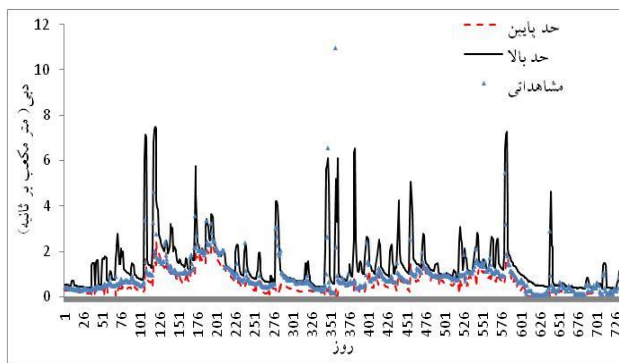
LEC بیشترین داده‌های خارج از دامنه اطمینان مربوط به داده‌های پیک می‌باشند. با مقایسه دو شکل (۷) و (۸) می‌توان مشاهده کرد روش LEC دامنه عدم قطعیت بزرگتری نسبت به روش رگرسیون چندک ارائه داده است. این در حالی است که در تحقیق Solomatine and Shrestha (2009) دامنه عدم قطعیت بدست آمده توسط روش رگرسیون چندک بزرگتر بود. از روش جنگل تصادفی نیز برای بررسی عدم قطعیت نتایج مدل هیدرولوژیکی استفاده شد. در روش جنگل تصادفی برای تعیین حداقل تعداد برگ در هر شاخه، با استفاده از آزمون و خطا تعداد برگ‌گی که در آن عملکرد مدل در برآورد دامنه عدم قطعیت بهترین نتیجه را ارائه داد به عنوان مقدار بهینه انتخاب گردید. در این تحقیق مشاهده شد چنانچه تعداد برگ کم باشد مدل بصورت موضعی کار کرده و نوسانات منحنی‌های حد بالا و حد پایین، زیاد می‌باشد و دامنه عدم قطعیت محاسباتی نیز بزرگ بوده است. در صورتی که با افزایش تعداد برگ از نوسانات



شکل ۸. دامنه عدم قطعیت محاسبه شده با استفاده از رگرسیون چندک



شکل ۹. دامنه عدم قطعیت محاسبه شده با استفاده از روش جنگل تصادفی شکل ۱۰. دامنه عدم قطعیت محاسبه شده با روش GLUE



جدول ۲. مقادیر معیارهای ارزیابی برای روش‌های مختلف برآورد عدم قطعیت

معیار آماری	رگرسیون چندک	LEC آموزش داده با ماشین بردار	LEC بدون آموزش	جنگل تصادفی	GLUE
PICP	۰/۸۶۶	۰/۹۶۷	۰/۹۵۱	۰/۸۱۲	۰/۶۹۳
MPI	۱/۱۶۴	۱/۳۰۷	۱/۱۹۹	۰/۸۴۲	۰/۹۴۹
ARIL	۱/۷۲۱	۲/۲۲۶	۲/۰۹۹	۱/۲۸۲	۱/۴۶۱

در روش جنگل تصادفی اگر چه مقدار دامنه عدم قطعیت برای داده‌های متوسط خیلی کم است اما با توجه به ورودی‌ها و فراوانی بیشتر دبی‌های نسبت به دبی‌های سیلابی بوده‌اند، مدل به نوعی توانسته حافظه بلند مدت موجود در داده‌ها را حفظ کند و در برآورد دامنه عدم قطعیت استفاده کند در این حالت تعداد بیشتری از داده‌های مشاهداتی در دامنه عدم قطعیت قرار گرفته‌اند. در بخش دیگری از تحقیق به منظور بررسی کارایی روش‌های مبتنی بر تجزیه و تحلیل باقیمانده‌ها (روش‌های ناپارامتریک) نسبت به روش پارامتریک، دامنه عدم قطعیت توسط روش GLUE برآورد گردید. این روش از زمان ارائه توسط Beven and Binley (1992) بطور گسترده‌ای در برآورد عدم قطعیت مدل‌های هیدرولوژیکی استفاده شده است و در آن تأثیر منابع عدم قطعیت مختلف در قالب عدم قطعیت پارامترها بررسی می‌شود. در این روش با نمونه‌برداری به تعداد زیاد از مجموعه پارامترها و اجرا کردن مدل با در نظر گرفتن معیار ناش ساتکلیف به عنوان تابع احتمال دامنه عدم قطعیت برآورد می‌شود. این روش بر خلاف روش‌ها قبل نیازی به مقادیر ورودی‌ها در گام‌های زمانی قبل نداشته و به نوعی فقط براساس مقادیر دبی دوره واسنجی اقدام به برآورد دامنه عدم قطعیت برای دوره صحت سنجی می‌کند. برای اجرای این روش بین کدهای روش GLUE و مدل هیدرولوژیکی HBV در نرم‌افزار متلب لینک برقرار گردید و با در نظر گرفتن توزیع پیشین یکنواخت برای تمام پارامترها، به تعداد ۱۵۰۰۰ مرتبه از مجموعه پارامترها نمونه‌برداری شد و مدل HBV با این مجموعه پارامترها اجرا گردید. در شکل (۱۰) دامنه عدم قطعیت برآورد شده توسط روش GLUE ارائه شده است. توجه به این شکل نشان می‌دهد در اغلب موارد داده‌های مشاهداتی نزدیک حد بالا هستند و حد پایین در برخی موارد مقداری نزدیک به صفر دارد و به نوعی می‌توان گفت حد پایین آن یک حالت بینابین روش LEC و رگرسیون چندک است. در این روش برای برخی داده‌های سیلابی حد بالای بزرگی برآورد شده است برخی داده‌های سیلابی که خارج از دامنه عدم قطعیت بدست آمده توسط روش‌های رگرسیون چندک و LEC قرار گرفته‌اند، در این روش در محدوده دامنه عدم قطعیت قرار گرفته‌اند.

در جدول (۲) نتایج معیارهای ارزیابی مربوط به روش‌های مختلف ارائه شده است. با توجه به جدول (۲) ملاحظه می‌شود کمترین مقادیر PICP مربوط به روش جنگل تصادفی و بیشترین مقدار مربوط به روش LEC است، در حالتی که آموزش داده شده باشد. با در نظر گرفتن مقدار حد اطمینان ۹۰ درصدی (هر قدر مقدار PICP به مقدار ۰/۹ نزدیکتر باشد

عملکرد آن روش بهتر است)، مقادیر PICP بدست آمده توسط روش‌های رگرسیون چندکی و LEC در حالتی که آموزش ندیده باشد بهتر هستند. کمترین مقادیر MPI مربوطه به روش جنگل تصادفی و بیشترین مقدار آن مربوط روش LEC در حالتی که آموزش دیده باشد، است. بنابراین با توجه به این معیار بهترین عملکرد مربوط به روش جنگل تصادفی و سپس روش GLUE است. (۲) در مقایسه با نتایج بدست آمده در تحقیق Solomatine and Shrestha (2009)، در این تحقیق دامنه عدم قطعیت محاسبه شده با روش LEC، تعداد بیشتری داده‌های مشاهداتی (مقدار شاخص PICP بزرگتر) را در بر می‌گیرد. در تحقیق Dogulu *et al* (2015) مشاهده کردند برای داده‌های متوسط و سیلابی دامنه عدم قطعیت و مقادیر پارامترهای PICP، MPI بدست آمده توسط رگرسیون چندک بیشتر از مقادیر بدست آمده توسط روش LEC بود که یک دلیل این مسئله را می‌توان در تعداد خوشه‌ها و همچنین نوع ورودی‌ها در نظر گرفته شده در روش LEC دانست.

به طور کلی نتایج بدست آمده در این قسمت نشان می‌دهد روش رگرسیون چندکی، سپس روش LEC بدون آموزش، بهترین عملکرد را در برآورد عدم قطعیت را داشته‌اند و بدترین عملکرد نیز توسط جنگل تصادفی بدست آمده است. در تحقیق Dogulu *et al* (2015) در برخی حوضه‌ها مدل رگرسیون چندک و در برخی از حوضه‌ها روش LEC عملکرد بهتری نسبت به یکدیگر داشته‌اند. در تحقیقات Lopez *et al* (2011); Weerts *et al* (2014) روش رگرسیون چندک عملکرد خوبی در برآورد دامنه عدم قطعیت داشته است. در مقایسه با روش‌های ناپارامتریک روش GLUE بدترین عملکرد را از نظر شاخص عمل PICP داشته است و از نظر معیارهای MPI و ARIL بهتر عمل کرده است. در مقایسه با تحقیق Solomatine and Shrestha (2009)، (روش GLUE دامنه عدم قطعیت کوچکتری نسبت به روش‌های دیگر ارائه داده بود) اگر چه دامنه عدم قطعیت بدست آمده بزرگتر است، اما همچنان از نظر تعداد داده‌های خارج از دامنه عدم قطعیت روش GLUE عملکرد مناسبی نداشت.

### نتیجه‌گیری

معمولاً بین نتایج همه مدل‌ها و مقادیر مشاهده شده تفاوت وجود دارد که عموماً عدم قطعیت خوانده می‌شود. موضوع شبیه‌سازی رواناب همراه با عدم قطعیت است که بسیاری از آن عدم قطعیت‌ها فقط به طور نسبی و تعداد محدودی از آن عدم قطعیت‌ها توسط ابزارهای آنالیزی موجود شناخته شده است. بعضی از این عدم قطعیت‌ها ریشه در زیربنای علمی دارند، در

ناپارامتریک روش پارامتریک GLUE از نظر معیار PICP بدترین عملکرد و از نظر معیارهای MPI و ARIL عملکرد بهتری داشته است. در تمام مدل‌ها بیشترین داده‌های خارج از دامنه عدم قطعیت مربوط به جریان‌های سیلابی بوده است که یک دلیل این مسئله را می‌توان در تعداد کم داده‌های سیلابی در دوره صحت‌سنجی و دلیل دیگر آن را می‌توان در آبی بودن و عدم وجود حافظه بلند مدت در داده‌های سیلابی دانست. نتایج کلی این تحقیق بیانگر عملکرد بهتر روش رگرسیون چندک در برآورد دامنه عدم قطعیت نسبت به سه روش دیگر مورد بررسی در این تحقیق است. در روش رگرسیون چندک تنها ورودی در نظر گرفته شده مقادیر دبی برآوردی است که در این حالت فرض می‌شود منابع مختلف عدم قطعیت از طریق مقادیر برآوردی بر روی خروجی مدل مؤثر می‌باشند. به عبارت دیگر در روش رگرسیون چندک فقط یک ورودی لازم بوده و از نظر ساختار نسبت به روش LEC که تعداد ورودی‌های بیشتر و همچنین مراحل بیشتری دارد، ساده‌تر است. در روش‌های مبتنی بر تجزیه و تحلیل باقیمانده‌ها داده‌های دوره واسنجی باید داده‌های حدی را در خود داشته باشد در غیر این صورت مدل اقدام به برون‌یابی می‌کند که در این حالت مقادیر خطای برآورد دامنه عدم قطعیت به شدت افزایش می‌یابد.

## REFERENCES

- AghaKouchak, A. and Habib, E. (2010). Application of a conceptual hydrologic model in teaching hydrologic processes. *Int. J. Eng. Educ.* 26, 963–973.
- Beven, K. and Binley, A. (1992). The future of distributed models: Model calibration and uncertainty prediction. *Hydrol. Process.*, 6, 279–298.
- Bezdek, J.C. (1974a). Numerical taxonomy with fuzzy sets. *Journal of Mathematical Biology*, 1: 57–71.
- Bezdek, J.C. (1974b). Cluster validity with fuzzy sets. *Journal of Cybernetics*, 3 (3), 58–72.
- Breiman, Leo. (2001). Random forests. *Mach. Learn.* 45 (1), 5–32.
- Chang, C. H., Yang J.C. and Tung, Y.K. (1993). Sensitivity and uncertainty analysis of a sediment transport models: a global approach. *Stochastic Hydrological Hydraulics*, 7 (4), 299–314.
- Dogulu, N., López López, P., Solomatine, D. P., Weerts, A. H. and Shrestha, D. L. (2015). Estimation of predictive hydrologic uncertainty using the quantile regression and UNEEC methods and their comparison on contrasting catchments. *Hydrol. Earth Syst. Sci.*, 19, 3181–3201.
- Evin, G., Thyer, M. Kavetski, D. McInerney D. and Kuczera, G. (2014). Comparison of joint versus postprocessor approaches for hydrological uncertainty estimation accounting for error autocorrelation and hetero-scedasticity. *Water Resour. Res.* 50 (3), 2350–2375.
- Fukuyama, Y. and Sugeno, M. (1989). A new method of choosing the number of clusters for the fuzzy c-means method. *Proceedings of Fifth Fuzzy Systems Symposium*, pp. 247–250 (in Japanese)
- Houska, T., Multsch, P., Kraft, H., Frede, G. and Breuer, L. (2014). Monte Carlo-Based calibration and uncertainty analysis of a coupled plant growth and hydrological model. *Biogeosciences*, 11, 2069–2082.
- Khu, S.T. and Werner, M.G.F. (2003). Reduction of monte-carlo simulation runs for uncertainty estimation in hydrological modeling. *Hydrology and Earth System Sciences*. 7 (5), 680–692.
- Koenker, R. (2005). *Quantile Regression*, Cambridge University Press.
- Koenker, R. and Bassett, J.r. (1978). *Regression Quantiles*, *Econometrica*, 1, 33–50.
- López López, P., Verkade, J.S., Weerts, A .H. and Solomatine, D.P. (2014). Alternative configurations of quantile regression for estimating predictive uncertainty in water level forecasts for the Upper Severn River: a comparison. *Hydrol. Earth Syst. Sci. Discuss.* 11 (4), 3811 – 3855.
- Malone, B.P., McBratney, A.B. and Minasny, B. (2011). Empirical estimates of uncertainty for mapping continuous depth functions of soil

حالی که دیگر عدم قطعیت‌ها مرتبط با عدم فهم کامل چگونگی اثر ماهیت پیچیده طبیعت می‌باشد. ارائه نتایج مدل بدون در نظر گرفتن عدم قطعیت آن فاقد ارزش برای تصمیم‌گیرندگان است. در اغلب تحقیقات برآورد دامنه عدم قطعیت با استفاده از روش‌های پارامتریک، دامنه عدم قطعیت برای یک منبع خطا برآورد می‌گردد. در حالی که با استفاده از روش‌های مبتنی بر تجزیه و تحلیل باقیمانده‌ها، عدم قطعیت ناشی از تمام منابع خطا که به نوعی بر خروجی مدل مؤثر هستند برآورد می‌گردد. بدین منظور در این تحقیق اقدام به برآورد دامنه عدم قطعیت خروجی مدل هیدرولوژیک HBV با استفاده از روش‌های مبتنی بر تجزیه و تحلیل مقادیر باقیمانده‌ها گردید. در ابتدا با استفاده روش تکامل تصادفی جوامع پارامترهای مدل HBV بهینه گردید و در گام بعد مدل برای دوره صحت‌سنجی اجرا گردید. پس از اجرای مدل مقادیر باقیمانده‌ها برای دوره واسنجی و دوره صحت‌سنجی محاسبه گردید و روش‌های مختلف برآورد عدم قطعیت اجرا گردید. نتایج نشان داد کمترین مقدار دامنه عدم قطعیت مربوط به روش جنگل تصادفی و بزرگترین دامنه عدم قطعیت مربوط به روش LEC می‌باشد. روش LEC بدون آموزش با هوش مصنوعی عملکرد بهتری نسبت به روش LEC آموزش داده شده با هوش مصنوعی، داشت. در مقایسه با روش‌های

- attributes. *Geoderma*, 160 (3–4), 614-626.
- Matott, L.S., Babendreier, J.E. and Purucker, S.T. (2009). Evaluating uncertainty in integrated environmental models: A review of concepts and tools. *Water Resources Research*, 45, W06421.
- Montanari, A. (2011). Uncertainty of Hydrological Predictions. In: Peter Wilderer (ed.) *Treatise on Water Science*, vol 2. pp. 459–478 Oxford: Academic Press.
- Rouhani, H. and Farahi Moghadam, M. (2014). Application of the Genetic Algorithm Technique for Optimization of the Hydrologic Tank and SIMHHYD Models' Parameters. *Journal Of Range and Watershed Management (Iranian Journal Of Natural Resources)*. 66(4), 521-533. In Farsi).
- Shrestha, D. L., and D. P. Solomatine (2006). Machine learning approaches for estimation of prediction interval for the model output , *Neural Networks* , 19 (2), 225 – 235, doi: 10.1016/j.neunet. 2006. 01. 012.
- Siebert, J. and Vis, M. J. P. (2012). Teaching hydrological modeling with a userfriendly catchment runoff-model software package, *Earth Syst. Sci.* 16, 3315-3325.
- Solomatine, D. P. and Siek, M. B. (2006). Modular learning models in forecasting natural phenomena. *Neural Networks*, 19(2), 215-224.
- Solomatine, D. P. and Shrestha, D. L. (2009). A novel method to estimate model uncertainty using machine learning techniques, *Water Re-sour. Res.*, 45, W00B11.
- Walker, W.E., Harremoës, P., Rotmans, J., Van der Sluis, J.P., Van Asselt, M.B.A., Janssen, P. and Kraye von Krauss M.P. (2003). Defining uncertainty a conceptual basis for uncertainty management in model-based decision support. *Integrated Assessment*, 4 (1), 5-17
- Weerts, A. H., Winsemius, H. C. and Verkade, J. S. (2011). Estimation of predictive hydrological uncertainty using quantile regression: examples from the National Flood Forecasting System (England and Wales), *Hydrol. Earth Syst. Sci.*, 15, 255–265.
- Xu, T., and Valocchi, A. J. (2015). Data-driven methods to improve base flow prediction of a regional groundwater model. *Computers & Geosciences*. 85(B), 124-13.