



دوره ۳۲، شماره ۳، شماره پیاپی ۱۲۴، پاییز ۱۳۹۸، صفحات ۷۷-۶۳
شناسه‌ی دیجیتال: 10.22092/wmej.2019.124804.1187

پژوهش‌های آبخیزداری

بررسی کارایی روش‌های نزدیک‌ترین همسایه و مبتنی بر خوشه‌بندی فازی در ترکیب خروجی مدل‌های آب‌سنجی

روژین دلپسند

دانشجوی کارشناسی ارشد آبخیزداری دانشگاه گنبد کاووس

ابوالحسن فتح‌آبادی*

(نویسنده‌ی مسئول)* استادیار گروه مرتع و آبخیزداری دانشگاه گنبد کاووس

حامد روحانی

استادیار گروه مرتع و آبخیزداری دانشگاه گنبد کاووس

سید مرتضی سیدیان

استادیار گروه مرتع و آبخیزداری دانشگاه گنبد کاووس

*رایانامه‌ی نویسنده‌ی مسئول: ahfathabadi@gmail.com

تاریخ دریافت: ۱۳ دی ۱۳۹۷ تاریخ پذیرش: ۲۸ خرداد ۱۳۹۸

چکیده

داده‌های ورودی و ساختار مدل‌ها معمولاً نقص‌هایی دارند و هیچ مدل منفردی را نمی‌توان پیدا کرد که عمل کرد آن در شبیه‌سازی جریان رود در تمام شرایط بهترین بوده و در خروجی آن بی‌قطعی نباشد. در این حالت با ترکیب مدل‌ها از مزیت‌های هر یک از مدل‌های منفرد برای ساختن مدلی که عمل کرد بهتری از هر یک از مدل‌های منفرد دارد بهره گرفته می‌شود. در این تحقیق، کارایی روش‌های نافرسانجه‌ی نزدیک‌ترین همسایه و خوشه‌بندی فازی نسبت به روش‌های ترکیب مدل (BGA (Bates Granger Averaging), GRA (Granger Ramanathan Averaging), BICA (Bayes Information Criterion), AICA (Akaike Information Criterion)، متوسط‌گیری با وزن‌های یکسان و روش لاسو در ترکیب خروجی مدل‌های آب‌شناختی یکپارچه GR5J، SimHyd، SACRAMENTO و SMAR بررسی شد. با کاربرد داده‌های ورودی بارش، دما، آب‌دهی و تبخیر تعرق هر یک از مدل‌های منفرد واسنجی، و روان‌آب خروجی حوضه‌ی کسلیان شهرستان پل سفید در ایستگاه ولیکن در مقیاس روزانه برآورد شد. سپس هر یک از روش‌های ترکیب مدل‌ها برای ترکیب نتایج خروجی هر یک از مدل‌های منفرد اجرا شد. نتایج نشان دادند که بهترین عمل کرد در دوره‌ی واسنجی در مدل‌های GR5J و SACRAMENTO، و در دوره‌ی اعتبارسنجی در مدل‌های SimHyd و GR5J بود. بهترین عمل کرد مدل‌های ترکیبی در دوره‌ی واسنجی در روش‌های لاسو و GRA بود که هر دو مشابه هم عمل کردند؛ اندازه‌های ضریب همبستگی، ضریب نش-ساتکلیف و RMSE آن‌ها به ترتیب ۰/۸۳، ۰/۶۹ و ۰/۲۴ بود. در دوره‌ی اعتبارسنجی برای روش‌های متوسط‌گیری با وزن‌های یکسان و روش BGA با اندازه‌های ضریب همبستگی، ضریب نش-ساتکلیف و RMSE به ترتیب ۰/۷۳، ۰/۲۷ و ۰/۵۲ بود. در دوره‌ی واسنجی عمل کرد روش نزدیک‌ترین همسایه بهتر از روش مبتنی بر خوشه‌بندی فازی بود، و بهترین عمل کرد هر دو مدل در ۲۰ همسایه به دست آمد. در دوره‌ی اعتبارسنجی عمل کرد روش مبتنی بر خوشه‌بندی فازی بهتر بود و عمل کرد هر دو مدل با افزایش تعداد همسایه بهتر شد.

واژگان کلیدی: بارش روان‌آب، کسلیان، ترکیب مدل، مدل GR5J، مدل SACRAMENTO

مقدمه

مدل‌ها برای پاسخ‌گویی به سوال‌های گوناگون در علوم مختلف از جمله مدیریت منابع آب گسترش داده شده‌اند. مدل‌های آب-شناختی نمایش ساده‌ی از فرآیندهای فیزیکی است که در حوضه اتفاق می‌افتد، و با دریافت ورودی‌هایی مثل بارش، دما و تبخیر و تعرق روان آب خروجی حوضه را شبیه‌سازی می‌کنند (لی و سانکاراسوبرامانیان ۲۰۱۲). به دلیل عواملی از قبیل نقص داده‌های به کار رفته برای شرایط اولیه‌ی مدل، خطاهای داده‌های ورودی، نقص ساختار مدل و ناتوانی در پیدا کردن سنج‌های بهینه‌ی مدل، همواره خروجی مدل‌های آب‌شناختی عدم قطعیت‌هایی دارد.

در مدل‌های آب‌شناختی به دلیل نداشتن شناخت کامل از سامانه‌ی آب‌شناختی واقعی، برای کاهش دادن عدم قطعیت خروجی مدل‌ها، می‌توان راه‌برد جایگزین چندین مدل و ترکیب خروجی آن‌ها را به کار گرفت (سینگ و همکاران ۲۰۱۴). روش‌های ترکیب مدل‌ها در هیدرولوژی را می‌توان به سه دسته تقسیم کرد. دسته‌ی اول بر اساس کاربرد نتایج خروجی هر یک از مدل‌های منفرد در گام‌های زمانی مختلف است. یکی از این گروه، روش تغییر مکان احتمالی است، که در آن در هر گام خروجی یکی از مدل‌های منفرد، خروجی مدل ترکیبی گرفته می‌شود (سی و همکاران ۲۰۰۰). در روی‌کردی مشابه تغییر مکان بر اساس تراز آب‌دهی بود (هیو و همکاران ۲۰۰۱). نتایج مقایسه‌ی شش روش ترکیبی شامل روش‌های ترکیبی مبتنی بر بهینه‌سازی وزن‌ها (متوسط‌گیری ساده و دو ساختار شبکه‌ی عصبی) و تغییر مکان (تغییر مکان احتمالی که مبتنی بر عمل‌کرد مدل در دوره‌ی زمانی قبل بود، و دو ساختار فازی مبتنی بر عمل‌کرد مدل در دوره‌ی زمانی قبل) در دو حوضه نشان داد که در حالتی که وضعیت جریان ایستا باشد و تغییرات زیاد نباشد روش شبکه‌ی عصبی، و در حالتی که وضعیت جریان متغیر باشد روش فازی عمل‌کرد بهتری دارد (آبراهارت و همکاران ۲۰۰۲). در دسته‌ی دیگر روش‌های ترکیب مدل‌ها بر اساس وزن‌دهی هر یک از مدل‌های منفرد است. از این دسته می‌توان به روش‌های متوسط‌گیری معمولی، متوسط‌گیری باتس و گرانجر (۱۹۶۹)، متوسط‌گیری مبتنی بر شاخص‌های آکاپک (بوکلند و همکاران ۱۹۹۷) و بیزی (اندرسون و بورنهان ۲۰۰۲)، و وزن‌های بهینه‌شده با روش کمینه‌ی مربع‌های معمولی (گرانجر و رامانتان ۱۹۸۴) اشاره کرد. نتایج ترکیب مدل‌ها با روش متوسط‌گیری ساده، وزن‌های بهینه‌شده‌ی کمینه‌ی مربع‌های معمولی و شبکه‌ی عصبی نشان داد که کاربرد مدل‌های ترکیبی باعث بهبود عمل‌کرد نسبت به هر یک از مدل‌های منفرد می‌شود (شمس‌الدین و همکاران ۱۹۹۷). نتایج دیک و روخت (۲۰۱۰) در ترکیب مدل‌های آب‌شناختی با کاربرد هفت روش وزن‌های یکسان، شاخص اطلاعات آکاپک، شاخص

اطلاعات بیزی، متوسط‌گیری Bates و Granger، متوسط‌گیری بیزی و متوسط‌گیری Mallows نشان داد که روش‌هایی که برای وزن‌ها محدودیتی نگذاشته‌اند نسبت به مدل‌هایی که برای وزن‌ها محدودیت قائل شده‌اند عمل‌کرد بهتری داشته‌اند. نتایج بررسی کارایی ۹ روش متوسط‌گیری در ترکیب نتایج چهار مدل آب‌شناختی نشان داد که سه روی‌کرد وزن‌دهی مبتنی بر Granger-Ramanathan و وزن‌های بهینه‌شده با روش تکامل تصادفی جوامع عمل‌کرد بهتری از روش‌های دیگر داشته‌اند (آرسنالت و همکاران ۲۰۱۵). نتایج ایسو و همکاران (۲۰۱۷) در بررسی کارایی دو روش میانگین‌گیری وزنی و GRA در دو حالت ممکن ترکیب خروجی مدل‌های نیواری (اتمسفری) در نقش ورودی مدل آب‌شناختی و ترکیب خروجی‌های مدل آب‌شناختی HSAMI برای هر ورودی، نشان داد که در حالتی که خروجی مدل آب‌شناختی ترکیب شدند عمل‌کرد مدل بهتر بود. در دسته‌ی دیگر از روش‌های ترکیب مدل‌ها، به جای مقدار برای هر گام زمانی تابع چگالی داده می‌شود. از این دسته می‌توان به متوسط‌گیری بیزی رفتاری و همکاران (۲۰۰۵)، هوتینگ و همکاران (۱۹۹۹) و ملاوز (هانسون ۲۰۰۷ و ۲۰۰۸) اشاره کرد. از این روش‌ها دان و همکاران (۲۰۰۷)، دیک و روخت (۲۰۱۰) و آرسنالت و همکاران (۲۰۱۵)، برای ترکیب نتایج مدل‌های آب‌شناختی به کار گرفتند. نتایج ژانگ و یانگ (۲۰۱۸) در ترکیب خروجی مدل‌های AWBM، GR5M، SimHyd، TRPWB، VWBM، YRWBM، VIC و شبکه‌ی عصبی مصنوعی با کاربرد روش‌های متوسط‌گیری بیزی، متوسط‌گیری با وزن‌های برابر و وزن‌های بهینه‌شده با الگوریتم ژنتیک نشان داد که الگوریتم ژنتیک عمل‌کرد بهتری از روش‌های متوسط‌گیری بیزی و متوسط‌گیری با وزن‌های برابر دارد. معمولاً مدل‌های منفرد می‌توانند برخی از فرآیندهای حوضه را به خوبی مدل‌سازی کند و دقت آن‌ها برای مدل‌سازی تمام فرآیندهای حوضه کافی نیست. در این حالت با ترکیب مدل‌ها می‌توان کارایی مدل‌ها را در برآورد اندازه‌های آب‌دهی بهبود داد. در روش‌هایی مانند متوسط‌گیری بیزی نیاز به برخی فرض‌ها مانند بهنجار بودن داده‌ها است، و در روش‌های وزن‌دهی نیاز به بهینه‌سازی وزن‌ها با توجه به تابع هدف انتخابی است. این در حالی است که معمولاً حساسیت تابع‌های هدف مثل نش-ساتکلیف یا RMSE به اندازه‌های خاصی از آب‌دهی بیش‌تر است، و کاربرد آن‌ها باعث ایجاد آرببی در نتیجه‌ی مدل می‌شود. در این تحقیق کارایی روش‌های نافرسانجه‌ی نزدیک‌ترین همسایه و خوشه‌بندی فازی که در آن‌ها تابعی خاصی بهینه نمی‌شود، نسبت به روش‌های فراسنجه‌ی بررسی شد. در مقایسه با تحقیقات دیگر، در این تحقیق روش لاسو و روش‌های نافرسانجه‌ی در ترکیب مدل‌ها به کار گرفته شد.

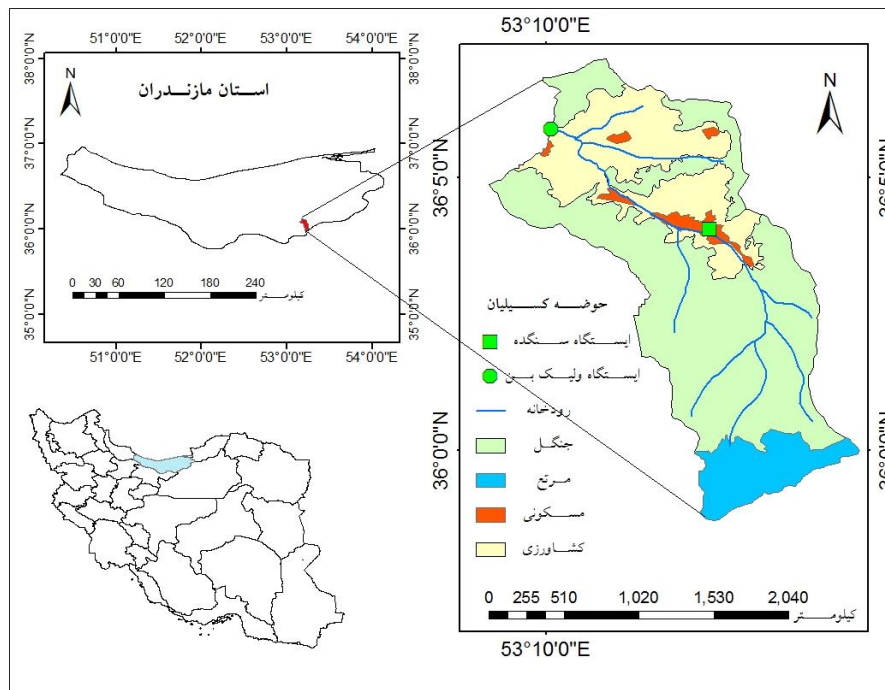
مواد و روش‌ها

منطقه‌ی بررسی شده

آبخیز کسلیان در $35^{\circ}58'30''$ تا $35^{\circ}07'15''$ عرض شمالی و $53^{\circ}08'44''$ تا $53^{\circ}15'42''$ طول شرقی در شهرستان پل سفید، از زیرحوضه‌های رودخانه‌ی تالار در استان مازندران، و مساحت آن $66/75$ کیلومتر مربع است. کم‌ترین، بیش‌ترین و متوسط ارتفاع منطقه به ترتیب 1120 ، 3349 و 1672 متر است. شبکه‌ی رودخانه‌ی حوضه که طول آبراه اصلی آن $16/5$ کیلومتر و میانگین شیب آن 13% است، حوضه را از جنوب به شمال غرب زه‌کشی می‌کند. متوسط بارندگی در این حوضه $783/4$ میلی‌متر و عمده‌ی بارش آن باران است. بخش‌های زیادی از این حوضه را گونه‌های جنگلی پوشانیده است. علاوه بر کاربری جنگلی،

کاربری‌های مرتعی، مسکونی و کشاورزی نیز در حوضه هست (شکل ۱) (میرعباسی نجف‌آبادی و همکاران ۲۰۱۲، سبزواری و همکاران ۲۰۰۸).

در این تحقیق به ترتیب داده‌های هواشناسی و آب‌دهی ایستگاه‌های باران‌سنجی سنگده و آب‌سنجی ولیک‌بین در مقیاس روزانه به کار گرفته شد. با توجه به نبود برخی از سنجه‌های اقلیمی در برخی از سال‌ها، داده‌های سال‌های 1387 تا 1392 برای مدل‌سازی به کار برده شد. برای واسنجی و اعتبارسنجی مدل‌های به‌کاررفته، داده‌های مشاهده‌ی به دو قسمت واسنجی (داده‌های روزانه‌ی سال‌های 1387 تا 1390) و اعتبارسنجی (داده‌های روزانه‌ی 1390 تا 1392) تقسیم کرده شد.



شکل ۱- موقعیت آبخیز کسلیان.

بخشی از آب ذخیره‌شده در مخزن تولید، تبخیر (ES: تبخیر-تعرق ذخیره‌ی) می‌شود. بخشی از آب مخزن تولید نفوذ می‌کند و مانند آب پایه به روان‌آب حاصل از بارش اضافه می‌شود تا آب‌دهی کل برای روندیابی (Pr) به‌دست آید (تراور و همکاران ۲۰۱۴).

برای اجرای مدل GR بسته‌ی airGR نرم‌افزار R به‌کارگرفته شد. در این بسته شش ساختار مختلف از مدل GR هست، که در این تحقیق ساختارهای GR5J بررسی شد. پس از معرفی داده‌های ورودی به مدل، سنجه‌های مدل بهینه‌ی گرفته شد. برای بهینه‌سازی سنجه‌های مدل روش Michel به‌کارگرفته شد (کرون و همکاران ۲۰۱۸).

روش تحقیق

مدل‌های بارش روان‌آب

مدل GR (Genie Rural)

این مدل شامل دو بخش مخزن تولید و مخزن روندیابی است. ورودی‌های مدل شامل بارش و تبخیر-تعرق روزانه است. اگر بارش بیش‌تر از تبخیر-تعرق باشد مقادیر بارش خالص (Pn) و تبخیر-تعرق خالص محاسبه می‌شود. از کل بارش خالص Pn، مقدار Ps آن مستقیمانه وارد مخزن تولید می‌شود، که بیشینه‌ی حجم ذخیره‌ی آن $X1$ (میلی‌متر) و حجم ذخیره‌ی واقعی آن S (میلی‌متر) است. مقدار باقی‌مانده‌ی Pn-Ps تبدیل به روان‌آب می‌شود.

مدل SimHyd (Simple Hydrology)

در این مدل ابتدا بارش روزانه باعث پر شدن ظرفیت گیرش گیاهی حوضه می‌شود. قسمتی از بارش باقی‌مانده که بیش‌تر از ظرفیت نفوذ خاک است تبدیل به روان‌آب سطحی می‌شود. بارش نفوذیافته ممکن است تبدیل به روان‌آب بین‌لایه‌یی شود، یا به ذخیره‌ی آب زیرزمینی یا رطوبت خاک افزوده شود. در این مدل ۷ سنجه‌ی ثابت با الگوریتم ژنتیک بهینه‌سازی شد (چیو و همکاران ۲۰۰۲ و چیو و سیریوارندا ۲۰۰۵).

مدل

SMAR (Soil Moisture Accounting and Routing)

این مدل شامل دو بخش ترازنامه‌ی آب و روندیابی است. در بخش ترازنامه‌ی آب فرض می‌شود که حوضه از چند لایه تشکیل شده است و هر لایه توانایی ذخیره‌ی ۲۵ میلی‌متر رطوبت را در حد رطوبت زراعی دارد (ممکن است عمق لایه‌ی پایین‌کم‌تر از ۲۵ میلی‌متر باشد). عمق تمام لایه‌ها Z میلی‌متر فرض می‌شود. در این مدل برای هر بازه‌ی زمانی، عمق توان تبخیر برابر با اندازه‌های گرفته‌شده قرار داده، یا با روش پنمن‌مانتیث برآورد می‌شود، و در این حالت اندازه‌های توان تبخیر به‌دست‌آمده (potential) در T ضرب می‌شود. در مدل فرض می‌شود که مقدار تبخیر از لایه‌ی سطحی متناسب با توان تبخیر تعرق است در حالی که مقدار تبخیر از دومین لایه (اگر اولین لایه‌ی خاک رطوبت نداشته باشد) متناسب است با توان تبخیر که در C ، که کم‌تر از یک است، ضرب شده است. با خشک‌شدن دومین لایه، مقدار تبخیر در لایه‌ی سوم در C^2 ضرب می‌شود. مقدار تبخیر برای سایر لایه‌ها هم به همین روش محاسبه می‌شود. زمانی که مقدار بارش (R) بیش‌تر از مقدار توان تبخیر (E_p) باشد قسمتی از بارش افزوده اولین بخش روان‌آب $r_1 = \dot{H} \times (R - E_p)$ را تشکیل می‌دهد. اگر مقدار باقی‌مانده‌ی بارش افزوده $(1 - H^*) \times X$ بیش‌تر از مقدار نفوذ باشد، بارش افزوده دومین قسمت از روان‌آب سطحی $r_2 = (1 - H^*) \times X - Y$ را تشکیل می‌دهد. در حالتی که مقدار بارش افزوده کم‌تر از ظرفیت نفوذ باشد، آب نفوذیافته باعث افزایش رطوبت خاک از اولین لایه تا آخرین حد ظرفیت زراعی می‌شود. اگر پس از افزایش رطوبت لایه‌های خاک تا حد زراعی، بارشی بماند، سومین بخش از روان‌آب r_3 در نظر گرفته می‌شود. کل روان‌آب تولید شده $r = r_1 + r_2 + r_3$ است. (تان و اوکنور ۱۹۹۶).

مدل SACRAMENTO

در این مدل فرض می‌شود که در حوضه دو لایه‌ی خاک هست. لایه‌ی بالایی با ضخامت کم‌تر برای نشان‌دادن رطوبت سطحی خاک و برگ‌آب به‌کارگرفته می‌شود. در حالی لایه‌ی دوم با ضخامت بیش‌تر برای نشان‌دادن لایه‌ی عمقی خاک و ذخیره‌ی آب زیرزمینی به‌کارگرفته می‌شود. لایه‌ی بالایی دو منطقه‌ی آب کششی و منطقه‌ی آب آزاد، و لایه‌ی زیرین سه منطقه‌ی آب کششی، منطقه‌ی آب آزاد اولیه، و منطقه‌ی آب آزاد اضافی دارد. در مخزن آب کششی لایه‌ی سطحی، که آب در بین ذره‌های خاک با کشش سطحی نگه داشته می‌شود، آب ممکن است تبخیر یا به مخزن ذخیره‌ی آب آزاد اضافه شود. آب منطقه‌ی آزاد لایه‌ی اول ممکن است تبخیر شود، عمقی به لایه‌ی زیرین نفوذ کند، یا جانبی باعث تولید جریان بین‌لایه‌یی شود. در قسمت آب کششی لایه‌ی زیرین ممکن است آب تبخیر، یا به ذخیره‌ی آب آزاد اضافه شود. آب ذخیره‌شده‌ی آزاد اولیه ممکن است تبدیل به جریان پایه شود یا به مخزن آب آزاد اضافی نفوذ کند. در این مدل کل حوضه به دو قسمت مناطق نفوذپذیر و مناطق نفوذناپذیر تقسیم می‌شود. دریاچه‌ها، رودخانه‌ها، پیاده‌روها و سایر مناطق نفوذناپذیر که مستقیمانه در ارتباط با شبکه‌ی جریان است منطقه‌ی نفوذناپذیر دانسته می‌شود. در این مدل روان‌آب نهایی خروجی حوضه شامل روان‌آب مستقیم از سطح‌های نفوذناپذیر، روان‌آب بین‌لایه‌یی از مخزن آب آزاد اولین لایه و جریان آب پایه منشأگرفته از منطقه‌ی آب آزاد اولیه و منطقه‌ی آب آزاد اضافی است. این مدل ۱۶ سنجه را برای شبیه‌سازی روان‌آب به‌کار می‌برد: پنج سنجه برای تعریف بزرگی اندازه‌ی مخزن‌های رطوبت خاک، ۳ سنجه برای محاسبه‌ی میزان روان‌آب خروجی جانبی، ۳ سنجه برای محاسبه‌ی میزان نشست آب از مخزن‌های بالایی به مخزن‌های پایینی، ۲ سنجه برای محاسبه‌ی روان‌آب مستقیم، و ۳ سنجه برای محاسبه‌ی هدررفت‌های سامانه (کوناث‌پووکا و الدهو ۲۰۱۹). برای اجرای مدل‌های SimHyd، SMAR، Sacramento بسته‌ی نرم‌افزاری rtl^1 (پوگر ۲۰۰۴) به‌کارگرفته شد.

روش‌های ترکیب مدل‌ها

متوسط‌گیری با وزن‌های برابر (EWA)

در این روش وزن‌ها با رابطه‌ی

$$\hat{\beta}_{EWA} = \left(\frac{1}{k}, \dots, \frac{1}{k} \right)$$

و خروجی مشاهده‌ی است. با فرض‌های مدل وایازی خطی، برآوردگر OLS را می‌توان بهترین برآوردگر خطی ناریب دانست. حتی اگر برخی از این فرضیه‌های تأییدشده برآورده نشود، ممکن است این روش بتواند بهترین اندازه‌های سنجها را که به نوعی نزدیک اندازه‌های بهینه است، پیدا کند.

روش لاسو (Lasso)

در این روش برای بهینه‌سازی سنجها تابع هدف رابطه (۱) به کارگرفته می‌شود.

$$\sum_{i=1}^n \left(y_i - \beta_0 - \sum_{j=1}^p \beta_j x_{ij} \right)^2 + \lambda \sum_{j=1}^p |\beta_j| \quad (3)$$

اگر λ بزرگ باشد تمام ضریبها صفر، و اگر λ صفر باشد برآورد کمینه‌ی مربعها به دست می‌آید. در این تحقیق مقدار بهینه‌ی λ با اعتبارسنجی متقابل، با در نظر گرفتن ۵ دسته انتخاب شد. برای اجرای این روش در این تحقیق بسته‌ی glmnet نرم‌افزار R به کاربرده شد (فریدمن و همکاران ۲۰۱۸).

روش نزدیک‌ترین همسایه ۳

مراحل اجرای روش نزدیک‌ترین همسایه با در نظر گرفتن تعداد K همسایه به روش زیر است:

در این روش در ابتدا بردارهای وضعیت (D_i) متغیرهای مستقل که با کاربرد آن‌ها مقادیر متغیر وابسته در زمان حال برآورد می‌شود، (در این تحقیق بردارهای وضعیت مقادیر آبدهی‌های برآوردشده با مدل‌های منفرد برای هر گام زمانی بود) و بردار خلف (A_i) اندازه‌های آبدهی که هدف برآورد آن است تعیین می‌شود.

در گام بعد فاصله‌ی اقلیدسی (E_i) میان بردار وضعیت در زمان حال و بردار متغیرهای مستقل (بردار وضعیت) در گام‌های زمانی گذشته (در داده‌های آموزشی) با رابطه‌ی (۴) تعیین می‌شود.

$$r_{im} = \left(\sum_{j=1}^d w_j (v_j - v_j')^2 \right)^{1/2} \quad (4)$$

محاسبه می‌شود. در واقع در این روش به نوعی از خروجی مدل‌های دیگر میانگین‌گیری می‌شود، و برای تعیین وزن‌ها داده‌های آموزشی به کارگرفته نمی‌شود.

مدل متوسط بیتس گرینجر (BGA)

در این روش وزن‌های نهایی با رابطه‌ی $\hat{\beta}_{BGA,i} = \frac{1/\hat{\sigma}_i^2}{\sum_{j=1}^k 1/\hat{\sigma}_j^2}$ محاسبه می‌شود، که در آن $\hat{\sigma}_i^2$ نشان دهنده‌ی واریانس خروجی مدل i است، و برابر با واریانس باقی مانده‌ها $\mathbf{Y}_t = \mathbf{X}_{i,t} \mathbf{e}_{i,t}$ برای دوره‌ی واسنجی است.

روش‌های میانگین‌گیری مبتنی بر معیارهای آکایک (AICA) و شوارتز (BICA)

بوکلند و همکاران (۱۹۹۷) و بورنهام و همکاران (۲۰۰۲) کاربرد وزن‌هایی به شکل زیر را پیشنهاد دادند.

$$\hat{\beta}_i = \frac{\exp(-I_i/2)}{\sum_{j=1}^k \exp(-I_j/2)} \quad (1)$$

معیار اطلاعاتی است که با نکویی برازش مدل در دوره‌ی I_i واسنجی و پیچیدگی هر مدل متناسب است، و با رابطه‌ی

$$I_i = 2 - \log(L_i) + q(p_i)$$

بیان می‌شود. در این رابطه L_i تابع بیشینه‌ی احتمال مدل i و $q(p_i)$ مقدار جریمه‌ی در نظر گرفته‌شده برای پیچیدگی مدل است.

روش Granger-Ramanathan (GRA)

روشی طبیعی برای در نظر گرفتن همبستگی بین مدل‌های مختلف کاربرد برآوردگر OLS^2 در مدل وایازی (رگرسیون) خطی است (گرانجر و رامانتان ۱۹۸۴). برآوردگر OLS برای بردار سنجها β در مدل رگرسیون خطی رابطه‌ی زیر است:

$$\hat{\beta}_{GRA} = (X^T X)^{-1} X^T Y \quad (2)$$

که در آن X و Y به ترتیب خروجی هر یک از مدل‌های منفرد

- 2- ordinary least square
- 3- k-nearest neighbor
- 4- feature vector
- 5- successor vector
- 6- Euclidean distance

FCM با در نظر گرفتن تابع هدف رابطه (۸) ماتریس X به C مجموعه‌ی (خوشه) هم‌پوشانی‌دار خوشه‌بندی می‌شود.

$$J(U, V : X) = \sum_{i=1}^c \sum_{k=1}^N (U_{ik})^m d^2(X_k, V_i) \quad (۸)$$

$$\sum_{i=1}^c U_{ik} = 1 \quad \forall k \in \{1, \dots, N\} \quad (۹)$$

$$0 < \sum_{k=1}^N U_{ik} < N \quad \forall i \in \{1, \dots, c\} \quad (۱۰)$$

در رابطه‌های بالا $u_{ik} \in [0, 1]$ نشان‌دهنده‌ی درجه‌ی عضویت k امین بردار ویژگی X_k در i امین خوشه‌ی فازی با مرکز V_i است. ماتریس U نشان‌دهنده‌ی میزان عضویت هر یک از بردارهای ورودی در هر یک از خوشه‌ها و $m \in [1, \infty]$ نشان‌دهنده‌ی توان برای هر عضویت فازی است. $d^2(X_k, V_i)$ فاصله‌ی k امین بردار ویژگی از مرکز خوشه V_i است. در این رویکرد بعد از محاسبه‌ی مقدار عضویت هر یک از بردارهای ورودی به خوشه‌های مختلف، این اندازه‌های عضویت بردارهای وضعیت دانسته، و مراحل روش‌های نزدیک‌ترین همسایه برای تعیین خروجی ترکیبی دنبال می‌شود. برای اجرای روش‌های نزدیک‌ترین همسایه و روش مبتنی بر FCM در نرم افزار متلب شناسه‌نویسی شد.

نتایج و بحث

در جدول ۱ اندازه‌های معیارهای ارزیابی برای دوره‌ی واسنجی و اعتبارسنجی مدل‌های مختلف آورده شده است. برای دوره‌ی واسنجی بهترین عمل کرد در مدل GR5J با اندازه‌های RMSE، و ضریب همبستگی و نش-ساتکلیف به ترتیب ۰/۲۵، ۰/۸۲ و ۰/۶۷ است. با توجه به اندازه‌های شاخص‌های ارزیابی، مدل SACRAMENTO در دوره‌ی واسنجی عمل کرد بهتری از مدل‌های SimHyd و SMAR داشت. برای دوره‌ی اعتبارسنجی مدل SimHyd با اندازه‌های RMSE، ضریب همبستگی و نش-ساتکلیف به ترتیب ۰/۲۷، ۰/۷۳ و ۰/۵ بهترین عمل کرد بهتری از دو مدل SMAR و SACRAMENTO داشت. در شکل‌های ۲ و ۳ آب‌نگار اندازه‌های آبدهی مشاهده‌ی و شبیه‌سازی شده

v_j ، v_j امین جزء بردار وضعیت در زمان j ، t : امین جزء بردار وضعیت در زمان حال (i) و w_j مقدار وزن مربوط به j امین جزء بردار وضعیت است. در گام بعد هر یک از بردارهای وضعیت در گام‌های زمانی قبیل (داده‌های آموزش) بر اساس فاصله‌ی اقلیدسی محاسبه‌شده با رابطه‌ی ۱۷ رتبه‌بندی شد، به‌طوری که بردارهای تاریخی با کم‌ترین فاصله در رتبه‌ی ۱ و با بیش‌ترین فاصله‌ی اقلیدسی در رتبه‌ی آخر بود. در مرحله‌ی بعد k بردار وضعیت با کوچک‌ترین رتبه انتخاب (K) نزدیک‌ترین همسایه) و با رابطه‌ی ۵ مقدار تابع کرنل^۶ برای هر یک از همسایه‌ها محاسبه شد.

$$k(j(i)) = \frac{1/j}{\sum_{j=1}^k 1/j} \quad (۵)$$

در این رابطه k تعداد همسایه‌ها و J مرتبه هر یک از همسایه‌ها است. در آخرین مرحله با کاربرد رابطه‌ی ۶ اندازه‌ی آبدهی برآورد می‌شود.

$$(x_i) = \sum_{j=1}^k k(j(i)) x_{j(i)} \quad (۶)$$

x_i مقدار بردار خلف پیش‌بینی شده برای زمان حال و x_i مقدار بردار خلف مربوط به همسایه با مرتبه‌ی J است.

روی‌کرد مبتنی بر خوشه‌بندی فازی (FCM)

این الگوریتم بر بهینه‌سازی تکراری تابع هدف فازی بنا شده است. فرض می‌شود که X_k نشان‌دهنده‌ی k امین بردار ویژگی در فضای n بعدی با محورهای y_1, \dots, y_n است. در این جا X_{ik} نشان‌دهنده‌ی درایه i در بردار ویژگی n بعدی X_k است. این بردارها را می‌توان با ماتریس $n \times N$ بعدی X نمایش داد (رابطه‌ی ۷).

$$X = \begin{bmatrix} x_{11} & \dots & x_{1N} \\ \dots & \dots & \dots \\ x_{n1} & \dots & x_{nN} \end{bmatrix} \quad (۷)$$

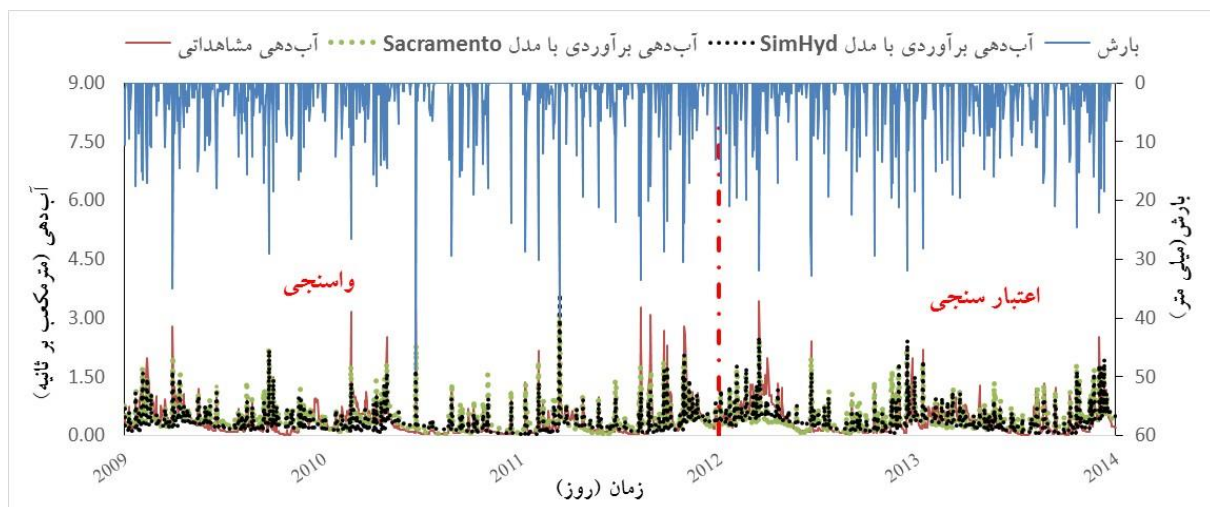
به‌علاوه، فرض می‌شود که بردار $V = (v_1, \dots, v_c)$ نشان‌دهنده‌ی مرکز هر یک از خوشه‌ها است. در الگوریتم

کند. در تحقیق روحانی و فراهی مقدم (۲۰۱۴) عمل کرد مدل SimHyd بهتر از مدل تانک بود و در دوره‌ی اعتبارسنجی توانست بهتر از دوره‌ی واسنجی اندازه‌های روان‌آب را برآورد کند. در تحقیق ژانگ و یانگ (۲۰۱۸) در مقایسه‌ی هشت مدل مختلف، عمل کرد مدل SimHyd فقط از مدل‌های VIC و VWBم بهتر بود، و در تحقیق ژانگ و همکاران (۲۰۱۳) عمل کرد مدل SACRAMENTO بهتر از مدل‌های GR۴J و SimHyd، و در تحقیق واز و همکاران (۲۰۱۰) عمل کرد همین مدل بهتر از مدل‌های AWBM و تانک بود.

با چهار مدل مختلف آورده شد. برای دوره‌ی واسنجی مدل SimHyd در اغلب موارد اندازه‌های آبدهی‌های اوج را کم‌تر از اندازه‌های مشاهده‌ی برآورد کرد. این مدل در برآورد اندازه‌های کمینه نیز عمل کرد مناسبی نداشت، و نتوانست نوسان‌های اندازه‌های آبدهی کمینه را به‌خوبی مدل‌سازی کند. مشابه همین را می‌توان در مدل SMAR مشاهده کرد. مدل SACRAMENTO هم در برآورد آبدهی اوج و هم در برآورد آبدهی‌های کمینه عمل کرد بهتری از دو مدل قبلی داشت، به‌طوری که این مدل توانست نوسان‌های آبدهی کمینه را به خوبی مدل‌سازی و شیب منحنی خشکیدگی را بهتر برآورد

جدول ۱- اندازه‌های معیارهای ارزیابی برای هر یک از مدل‌های منفرد در دوره‌ی آموزش و اعتبارسنجی.

GR5J	SimHyd	Sacramento	SMAR		
۰/۸۲	۰/۷۵	۰/۸۰	۰/۷۴	ضریب همبستگی	واسنجی
۰/۲۵	۰/۳۱	۰/۲۶	۰/۳۲	RMSE	
۰/۶۷	۰/۵۱	۰/۶۴	۰/۴۷	نش-ساتکلیف	
۰/۷۱	۰/۷۳	۰/۶۸	۰/۶۹	ضریب همبستگی	اعتبارسنجی
۰/۲۹	۰/۲۷	۰/۲۹	۰/۲۹	RMSE	
۰/۴۶	۰/۵۰	۰/۴۳	۰/۴۳	نش-ساتکلیف	



شکل ۲- آبنگار مشاهده‌ی و شبیه‌سازی شده با روش‌های Sacramento و SimHyd.

مدل‌ها آورده شد. در روش BGA تفاوت خیلی کمی بین وزن‌های اختصاص داده‌شده به مدل‌های مختلف هست. در این روش ضریب اختصاص داده‌شده به هر مدل متناسب با مقدار ضریب نش-ساتکلیف هر یک از مدل‌های منفرد در دوره‌ی واسنجی است. در حالی که در تحقیق ژانگ و یانگ (۲۰۱۸) وزن‌های به‌دست‌آمده با الگوریتم ژنتیک و متوسط‌گیری بیزی متناسب خطی با اندازه‌های نش-ساتکلیف به‌دست‌آمده در

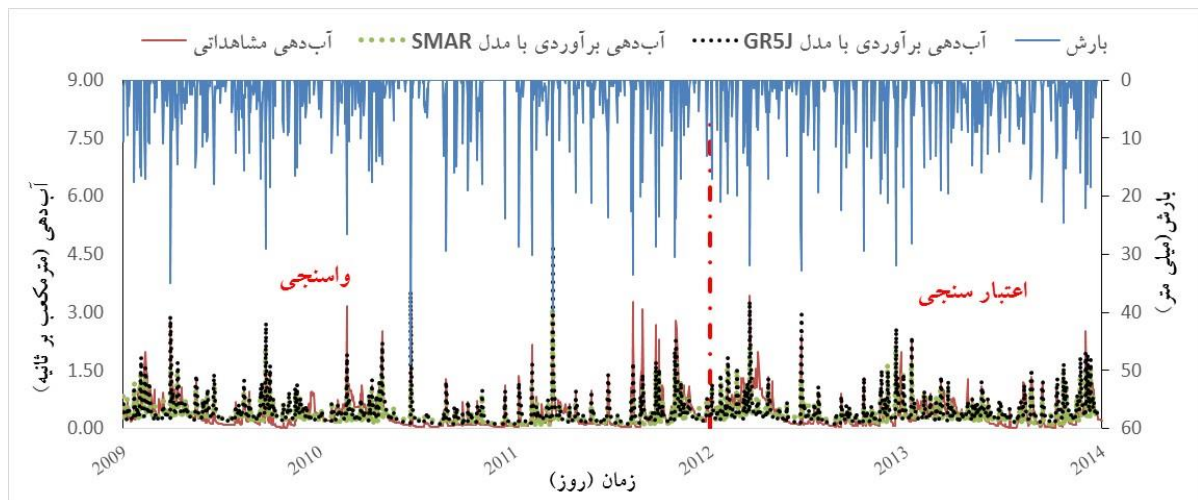
عمل کرد مدل GR5J از نظر شبیه‌سازی آبدهی‌های اوج، آبدهی‌های کمینه و شیب منحنی خشکیدگی بهتر از سه مدل دیگر بود (شکل‌های ۲ و ۳). در دوره‌ی اعتبارسنجی عمل کرد مدل SimHyd بهتر از سایر مدل‌ها به‌خصوص در شبیه‌سازی داده‌های اوج ابتدای دوره بود. پس از اجرای هر یک از مدل‌های منفرد، مدل‌های ترکیبی اجرا شد. در جدول ۲ اندازه‌های ضریب‌های ۶ روش ترکیب

بررسی کارایی روش‌های نزدیک‌ترین همسایه و مبتنی...

دوره‌ی آموزش نبود. در روش‌های AICA و BICA با توجه به عمل‌کرد بهتر مدل GR5J و تعداد سنج‌های کم‌تر آن، این مدل بیشینه‌ی وزن ممکن (۱) را داشت. در تحقیقات دیک و روخت (۲۰۱۰) و آرسنالت و همکاران (۲۰۱۵) نیز این دو روش به یک مدل وزن ۱ و سایر مدل‌ها وزن صفر دادند. اختلاف بین کم‌ترین و بیش‌ترین وزن داده‌شده در روش GRA بیش‌تر از روش‌های دیگر بود، و در این روش نیز بیش‌ترین وزن به مدل GR5J داده شد. در روش لاسو با توجه به تعداد داده‌ی مشاهده‌ی و بودن همبستگی بین خروجی مدل‌های مختلف، اندازه‌های ضریب‌های برآوردی اختلاف کمی با ضریب‌های محاسبه‌ی با روش GRA دارد.

در جدول ۳ اندازه‌های شاخص‌های ارزیابی برای خروجی مدل‌های ترکیبی برای داده‌های واسنجی و اعتبارسنجی آورده شد. برای داده‌های دوره‌ی واسنجی کم‌ترین مقدار RMSE و بیش‌ترین اندازه‌های نش-ساتکلیف و ضریب همبستگی در روش‌های لاسو و GRA، و بدترین عمل‌کرد نیز در روش متوسط‌گیری با وزن‌های برابر بود. برای داده‌های اعتبارسنجی با توجه به اندازه‌های شاخص‌های ارزیابی بهترین عمل‌کرد در

روش‌های BGA و متوسط‌گیری با وزن‌های برابر، و بدترین عمل‌کرد نیز در روش‌های AICA و BICA بود. یک دلیل این مسئله را می‌توان در عمل‌کرد متفاوت مدل SimHyd در دوره‌ی واسنجی و ارزیابی دانست. اگرچه عمل‌کرد روش‌های AICA، GRA و BICA در دوره‌ی واسنجی بهتر از مدل‌های BGA و میانگین‌گیری با وزن‌های یکسان بود، اما همین عمل‌کرد متفاوت مدل SimHyd در دوره‌ی آموزش و اعتبارسنجی منجر به آن شد. دیک و روخت (۲۰۱۰) نیز در بی‌تشابهی وزن‌های به‌دست‌آمده با بهترین روش‌های ترکیبی با اندازه‌های وزن‌های بهینه‌ی پیدا شده برای داده‌های اعتبارسنجی همین را مشاهده کردند. در تحقیق آرسنالت و همکاران (۲۰۱۵) عمل‌کرد روش میانگین‌گیری با وزن‌های برابر بهتر از دو روش BICA و AICA، و عمل‌کرد روش BGA بدترین بود. در تحقیق دیک و روخت (۲۰۱۰) عمل‌کرد روش GRA بهتر از روش‌های وزن‌های یکسان، AICA، BGA و BICA بود، و بدترین عمل‌کرد نیز در روش‌های وزن‌های برابر و BGA بود. در تحقیق دوبارکو و همکاران (۲۰۱۷) نیز عمل‌کرد روش GRA بهتر از روش BGA بود.



شکل ۳- آبنگار مشاهده‌ی و شبیه‌سازی شده با روش‌های GR5J و SMAR.

جدول ۲- وزن‌های به‌دست‌آمده برای هر یک از مدل‌های منفرد با روش‌های ترکیبی مختلف.

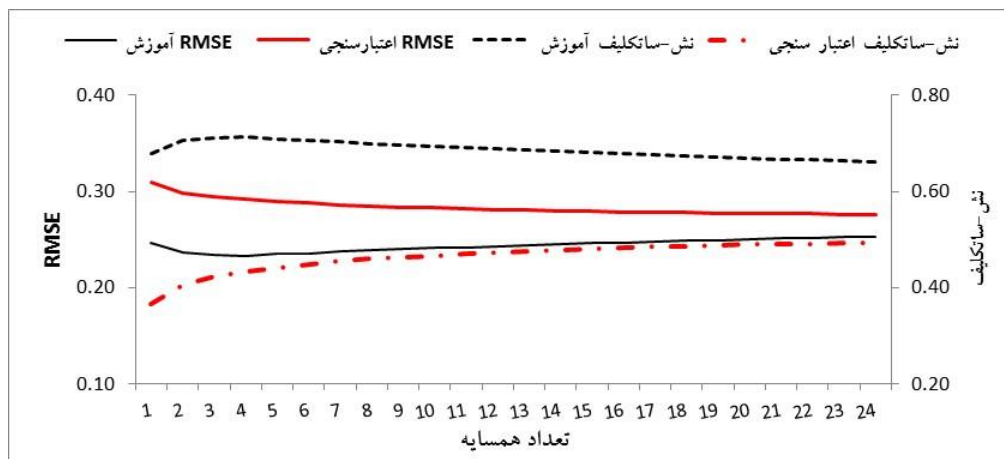
GR5J	SimHyd	Sacramento	SMAR	
۰/۲۵۰	۰/۲۵۰	۰/۲۵۰	۰/۲۵۰	EWA
۰/۲۹۵	۰/۲۲۲	۰/۲۷۲	۰/۲۱۰	BGA
۱/۰۰۰	۰/۰۰۰	۰/۰۰۰	۰/۰۰۰	AICA
۱/۰۰۰	۰/۰۰۰	۰/۰۰۰	۰/۰۰۰	BICA
۰/۱۶۶	۰/۴۳۵	۰/۴۴۶	-۰/۵۲۹	GRA
۰/۱۶۳۸	۰/۲۸۵	۰/۴۰۳	-۰/۳۱۶	Lasso
۰/۱۶۷۸	۱/۳۵۰	-۰/۲۳۸	-۰/۹۰۱	بهینه‌شده برای اعتبارسنجی

جدول ۳- اندازه‌های معیارهای ارزیابی مدل‌های ترکیبی برای داده‌های آموزش و اعتبارسنجی.

EWA	BGA	AICA	BICA	GRA	لاسو		
۰/۸۱	۰/۸۱	۰/۸۲	۰/۸۲	۰/۸۳	۰/۸۳	ضریب همبستگی	واسنجی
۰/۲۶	۰/۲۵	۰/۲۵	۰/۲۵	۰/۲۴	۰/۲۴	RMSE	
۰/۶۵	۰/۶۶	۰/۶۷	۰/۶۷	۰/۶۹	۰/۶۹	نش-ساتکلیف	
۰/۷۳	۰/۷۳	۰/۷۱	۰/۷۱	۰/۷۳	۰/۷۳	ضریب همبستگی	اعتبارسنجی
۰/۲۷	۰/۲۷	۰/۲۸	۰/۲۸	۰/۲۸	۰/۲۸	RMSE	
۰/۵۲	۰/۵۲	۰/۴۷	۰/۴۷	۰/۴۸	۰/۴۹	نش-ساتکلیف	

و همکاران (۲۰۱۵)، دوبارکو و همکاران (۲۰۱۷)، و محمد و همکاران (۲۰۱۸) نیز در مقایسه‌ی روش‌های مختلف ترکیب مدل‌ها مشاهده کردند که کاربرد مدل‌های ترکیبی لزوماً منجر به مدلی که عمل کرد آن بهتر از هر یک از مدل‌های منفرد است نمی‌شود. برای هر دو دوره‌ی واسنجی و اعتبارسنجی عمل کرد هر یک از مدل‌های ترکیبی بهتر از عمل کرد بدترین مدل منفرد بود، به آن معنی که خروجی مدل ترکیبی در بدترین حالت بهتر از خروجی بدترین مدل منفرد است.

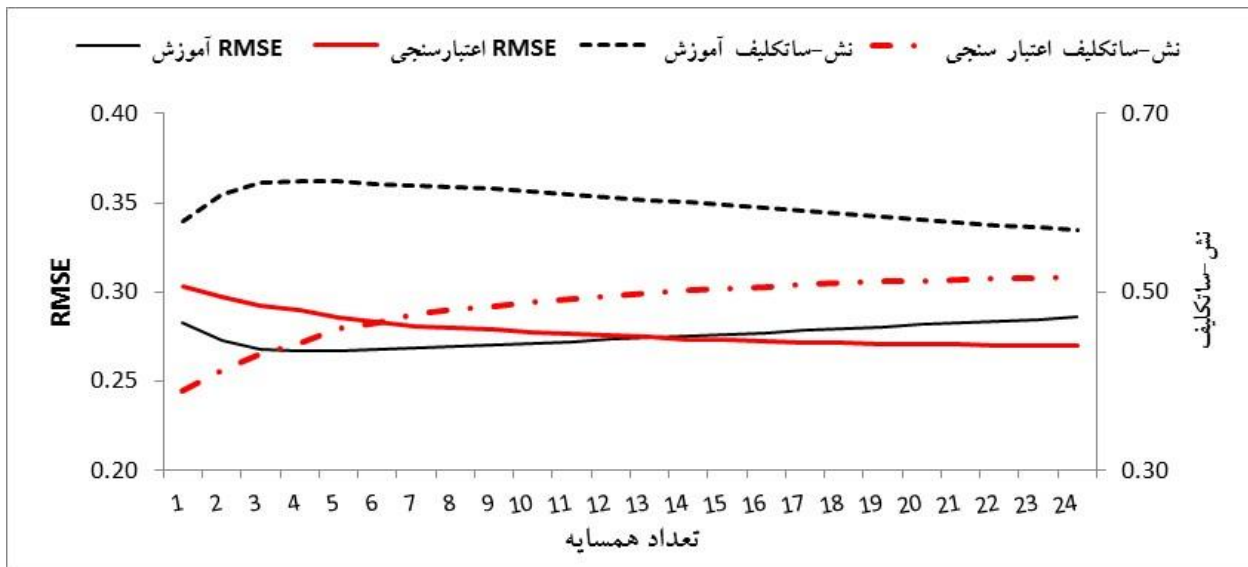
برای داده‌های دوره‌ی واسنجی عمل کرد روش‌های ACIA، BCIA و GRA مشابه و کمی بهتر از مدل GR5J که بهترین عمل کرد را داشت، بود و دو روش متوسط‌گیری با وزن‌های برابر و BGA نتوانست باعث بهبود عمل کرد نسبت به بهترین مدل‌های منفرد شود (جدول ۱ و ۳). در دوره‌ی اعتبارسنجی فقط برآورد دو مدل متوسط‌گیری با وزن‌های یکسان و BGA دقیق‌تر از مدل SimHyd بود، که بهتر از سایر مدل‌های منفرد اندازه‌های آب‌دهی را شبیه‌سازی کرد. دیک و روخت (۲۰۱۰)، آرسنالت



شکل ۴- اندازه‌های معیارهای ارزیابی برای تعداد همسایه‌های مختلف در روش نزدیک‌ترین همسایه.

در روش نزدیک‌ترین همسایه مسئله‌ی اصلی تعیین تعداد همسایه است. در این روش مدل برای تعداد همسایه‌های متفاوت اجرا شد، و نتایج آن برای معیارهای ارزیابی نش-ساتکلیف و RMSE در شکل ۴ دیده می‌شود. برای داده‌های دوره‌ی واسنجی، کمترین مقدار RMSE (۰/۲۳) و بیشترین مقدار نش-ساتکلیف (۰/۷۱) برای تعداد همسایه به دست آمد، و با افزایش تعداد همسایه عملکرد مدل بدتر شد (شکل ۴). برای این تعداد همسایه اندازه‌های RMSE، نش-ساتکلیف و ضریب همبستگی برای داده‌های اعتبارسنجی به ترتیب ۰/۲۹، ۰/۴۳ و ۰/۷۰ به دست آمد. برای داده‌های اعتبارسنجی، با افزایش تعداد همسایه تا ۱۰۰ روند

صعودی نش-ساتکلیف و نزولی RMSE شیب زیادی داشت و در بیش‌تر از این تعداد شیب تغییرات دو معیار ارزیابی کم‌تر شد (شکل ۴). در این حالت برای تعداد همسایه‌ی برابر با ۱۰۰ اندازه‌های RMSE، نش-ساتکلیف و ضریب همبستگی به ترتیب ۰/۲۸، ۰/۴۹ و ۰/۷۱ به دست آمد. با مقایسه‌ی این اندازه‌ها با اندازه‌های به دست آمده ملاحظه می‌شود که در دوره‌ی واسنجی این مدل فقط از روش‌های میانگین‌گیری با وزن‌های برابر و BGA عمل کرد بهتری داشت، و برای دوره‌ی اعتبارسنجی این روش بعد از روش‌های میانگین‌گیری با وزن‌های برابر و BGA عمل کرد بهتری از روش‌های دیگر داشت.



شکل ۵- اندازه‌های معیارهای ارزیابی برای تعداد همسایه‌های مختلف در روش خوشه‌بندی فازی (FCM).

نتیجه‌گیری

در این تحقیق ۸ روش ترکیب مدل‌ها در شبیه‌سازی فرآیند بارش روان‌آب در ایستگاه ولیکن حوضه‌ی کسلیان بررسی شد. روان‌آب خروجی حوضه با مدل‌های SACRAMENTO، SimHyd و GR5J شبیه‌سازی شد. نتایج نشان داد که بهترین عمل‌کرد در دوره‌ی واسنجی در مدل‌های GR5J و SACRAMENTO و در دوره‌ی اعتبارسنجی در مدل‌های SimHyd و GR5J بود. مدل‌های ترکیبی پس از اجرای مدل‌های منفرد اجرا شد. با کاربرد داده‌های واسنجی ضریب‌های هر یک از مدل‌های منفرد برآورد شد، و با کاربرد ضریب‌های برآوردشده‌ی اندازه‌های آب‌دهی برای دوره‌ی واسنجی برآورد شد. نتایج نشان داد که بهترین عمل‌کرد مدل‌های ترکیبی در دوره‌ی واسنجی در روش‌های لاسو و GRA و در دوره‌ی اعتبارسنجی در روش‌های وزن‌های یکسان و BGA به‌دست آمد. عمل‌کرد ضعیف‌تر مدل‌های لاسو و GRA در دوره‌ی واسنجی را می‌توان در عمل‌کرد متفاوت مدل‌های به‌کاررفته در دوره‌ی واسنجی و اعتبارسنجی دید، به طوری که عمل‌کرد مدل SimHyd اگر چه در دوره‌ی واسنجی مناسب نبود، اما در دوره‌ی اعتبارسنجی بهترین بود. همین کارکرد را می‌توان برای روش‌های نزدیک‌ترین همسایه مشاهده کرد، به طوری که در هر دو روش برای داده‌های واسنجی بهترین عمل‌کرد برای تعداد همسایه ۲۰ به‌دست آمد. در بیش‌تر از این تعداد همسایه، عمل‌کرد مدل بدتر شد، در حالی که برای داده‌های دوره‌ی اعتبارسنجی با افزایش تعداد همسایه‌ها عمل‌کرد مدل بهتر شد و ۱۰۰ همسایه بهترین دانسته

برای انتخاب‌کردن تعداد بهینه‌ی از خوشه‌ها در روش مبتنی بر FCM معیارهای V_{pe}^9 ، V_{pc}^8 ، V_{fs} (فوکویاما و سوگونو ۱۹۸۹) و V_{XB} (ژی و وینی ۱۹۹۱) به‌کارگرفته شد. نتایج نشان داد که کم‌ترین اندازه‌ی شاخص‌های V_{pe} ، V_{XB} و V_{fs} و بیش‌ترین اندازه‌های شاخص V_{pc} در حالتی بود که تعداد خوشه‌ها ۶ است، بنابراین، این تعداد خوشه بهینه دانسته شد. همانند روش نزدیک‌ترین همسایه، تعداد بهینه‌ی همسایه که عمل‌کرد مدل با آن بهترین بود نیز تعیین شد. تعداد همسایه بین ۵ تا ۱۲۰ در نظر گرفته شد (شکل ۵). برای داده‌های واسنجی همانند روش نزدیک‌ترین همسایه بهترین نتایج در حالتی بود که تعداد همسایه‌ها ۲۰ است. در این تعداد همسایه اندازه‌های RMSE، ضریب همبستگی و نش-ساتکلیف به ترتیب ۰/۲۷، ۰/۷۹ و ۰/۶۲ به‌دست آمد. برای داده‌های اعتبارسنجی همانند روش نزدیک‌ترین همسایه شیب تغییرات اندازه‌های RMSE و نش-ساتکلیف تا حدود ۱۰۰ همسایه زیاد بود و از این تعداد بیش‌تر شیب روند تغییرات کاهش می‌یافت (شکل ۵). در ۱۰۰ همسایه اندازه‌های معیارهای ارزیابی RMSE، ضریب همبستگی و نش-ساتکلیف به ترتیب ۰/۲۷، ۰/۷۲ و ۰/۵۱ به‌دست آمد. این روش در رتبه‌ی پس از روش‌های میانگین‌گیری با وزن‌های یکسان و BGA بود. برای دوره‌ی واسنجی، به‌ازای تعداد همسایه‌ی مشابه، عمل‌کرد روش نزدیک‌ترین همسایه، و برای دوره‌ی اعتبارسنجی عمل‌کرد روش مبتنی بر FCM بهتر بود.

8- partition coefficient

9- partition entropy

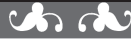
مبتنی بر خوشه‌بندی فازی بود. نتایج این تحقیق نشان می‌دهد که روش‌های مبتنی بر بهینه‌سازی لاسو، نزدیک‌ترین همسایه و FCM باعث می‌شود که عمل‌کرد بهتر از روش GRA شود، که در تحقیقات دیگر بهترین مدل شناخته شده بود.

شد. این اختلاف همسایه برای داده‌های آموزش و اعتبارسنجی را می‌توان در عمل‌کرد متفاوت مدل‌ها در دوره‌ی واسنجی و اعتبارسنجی دانست. بهترین عمل‌کرد در دوره‌ی واسنجی، در روش نزدیک‌ترین همسایه و در دوره‌ی اعتبارسنجی در روش

- Arsenault R, Gatién P, Renaud B, Brissette F, Martel JL. 2015. A comparative analysis of 9 multi-model averaging approaches in hydrological continuous streamflow simulation. *Journal of Hydrology*. 529(3): 754–767.
- Abrahart RJ, See L. 2002. Multi-model data fusion for river flow forecasting: an evaluation of six alternative methods based on two contrasting catchments. *Hydrology and Earth System Sciences*. 6(4): 655–670.
- Bates JM, Granger C. 1969. The combination of forecasts. *Journal of the Operational Research Society*. 20(4): 451–468.
- Bezdek JC. 1974a. Cluster validity with fuzzy sets. *J.Cybernet*. 3: 58–73.
- Bezdek JC. 1974b. Numerical taxonomy with fuzzy sets. *J.Math. Biol*. 1: 57–71.
- Buckland ST, Burnham KP, Augustin NH. 1997. Model selection: an integral part of inference. *Biometrics*. 53: 603–618.
- Burnham KP, Anderson DR. 2004. Multimodel inference: understanding AIC and BIC in model selection. *Sociological Methods & Research*. 33(2): 261–304.
- Borhani Darian AR, Farahmandfar Z. 2011. Calibration of rainfall-runoff models using MBO algorithm, *The Iranian Society of Irrigation and Water Engineering*. 1(4): 60–71.
- Duan Q, Ajami NK, Gao X, Sorooshian S. 2007. Multi-model ensemble hydrologic prediction using Bayesian model averaging. *Adv. Water Resour*. 30: 1371–1386.
- Chiew FHS, Peel MC, Western AW. 2002. Application and testing of the simple rainfall runoff model SimHyd, *Mathematical models of small watershed hydrology and application* (Eds: V.P.Singh and D.K.Frevert), Water Resources Publication, Littleton, Colorado, USA. pp. 335–367.
- Chiew FHS, Siriwardena L. 2005. Estimation of SimHyd parameter values for application in ungauged catchments, In Zerger A, and Argent RM, (eds) MODSIM 2005 International Congress on Modelling and Simulation, Modelling and Simulation Society of Australia and New Zealand.
- Coron L, Perrin C, Delaigue O, Thirel G, Michel C, Andréassian V, Bourgin F, Brigode P, Le Moine N, Mathevet T, Mouelhi S, Oudin L, Pushpalatha R, Valéry A, Coron L, Perrin C, Delaigue O, Thirel G, Michel C, Andréassian V, Bourgin F, Brigode P, Le Moine N, Mathevet T, Mouelhi S, Oudin L, Pushpalatha R, Valéry A. 2018. airGR: Suite of GR Hydrological Models for Precipitation-Runoff Modelling. R package.
- Diks CG, Vrugt JA. 2010. Comparison of point forecast accuracy of model averaging methods in hydrologic applications. *Stochastic Environmental Research and Risk Assessment*. 24 (6): 809–820.
- Dobarco MR, Arrouays D, Lagacherie P, Ciampalini R, Saby NPA. 2017. Prediction of topsoil texture for Region Centre (France) applying model ensemble methods. *Geoderma*. 298 (15): 67–77.
- Essou GRC, Brissette F, Picher PL. 2017. Impacts of combining reanalyses and weather station data on the accuracy of discharge modelling. *Journal of Hydrology*. 545: 120–131.
- Friedman J, Hastie T, Tibshirani R, Simon N, Narasimhan B, Qian j. 2018. glmnet: Lasso and Elastic-Net Regularized Generalized Linear Models. R package.
- Granger CW, Ramanathan R. 1984. Improved methods of combining forecasts. *Journal of Forecasting*. 3(2): 197–204.
- Granger CW, Newbold P. 1977. Identification of two-way causal systems. In *Frontiers of quantitative economics* (Vol. 3). North-Holland Amsterdam.

- Gu L, P. Hanson J, Post WM, Kaiser DP, Yang B, Nemani R, Meyers T. 2008. The 2007 eastern US spring freeze: increased cold damage in a warming world?. *AIBS Bulletin*.58(3): 253–262.
- Goodarzi MR, Zahabiyon B, M 18. Bavani B, Kamal RA. 1391. Performance comparison of three hydrological models SWAT, IHACRES and SimHyd for the runoff simulation of Ghare-sou basin. *Water and Irrigation Management*. 2(1): 25–40. (In Persian).
- Hu TS, Lam KC, Ng ST. 2001. River flow time series prediction with a range dependent neural network. *Hydrological Sciences Journal*. 46: 729–745.
- Hoeting JA, Madigan D, Raftery AE, Volinsky CT. 1999. Bayesian model averaging: a tutorial. *Statistical science*. 14(4): 382–417.
- Hansen BE. 2008. Least-squares forecast averaging. *J. Econom*. 146 (2): 342–350.
- Kachroo RK Liang GC. 1992. River flow forecasting. Part 2. Algebraic development of linear modelling techniques, *Journal of Hydrology*. 133:17–40.
- Kamal AR, Masah Bovani AR. 1389. Climate Change and Variability Impact in Basin's Runoff with Interference of Two Hydrology Models Uncertainty. *Journal of Water and Soil*. 24(5): 920–931. (In Persian).
- Kunnath-Poovakka A, Eldho TI. 2019. A comparative study of conceptual rainfall-runoff models GR4J, AWBM and Sacramento at catchments in the upper Godavari river basin, India. *J. Earth Syst. Sci*. 128(2):1-15.
- Li W, Sankarasubramanian A. 2012. Reducing hydrologic model uncertainty in monthly streamflow predictions using multimodel combination. *Water Resources Research*. 48(12).
- Mirabbasi Najafabadi R, Dinpazhoh Y, Fakheri-Fard A. 2012. Rainfall runoff modelling using the principle of maximum entropy (Case study: Kasilian Watershed). *Journal of Water and Soil Science*. 15(58): 39–51. (In Persian).
- Muhammad A, Stadnyk TA, Unduche F, Coulibaly P. 2018. Multi-Model approaches for improving seasonal ensemble streamflow prediction scheme with various statistical post-processing techniques in the Canadian Prairie Region. *Water*. 10(11).
- Parrish MA, Moradkhani H, DeChant CM. 2012. Toward reduction of model uncertainty: Integration of Bayesian model averaging and data assimilation. *Water Resources Research*. 48(3).
- Pogder G. 2004. *Rainfall Runoff Library User Manual*. CRC for Catchment Hydrology.
- Raftery AE, Gneiting T, Bakabdaoui F. 2005. Using Bayesian model averaging to calibrate forecast ensembles. *Mon. Weather. Rev*. 133 (5): 1155–1174.
- Rouhni H, Farahi Moghadam M. 2014. Application of the genetic algorithm technique for optimization of the hydrologic tank and SIM-HHYD models, parameters. *Journal of Range and Watershed Management*. 66(4): 521–533. (In Persian).
- Sabzevari T, Ardakanian R, Shamsayi A, Talebi A. 2008. Predicting the Flood Hydrographs of Ungauged Watersheds Using the HEC-HMS and GIS. *Water Engineering*. 2(4):1–12. (In Persian).
- Singh H, Sankarasubramanian A. 2014. Systematic uncertainty reduction strategies for developing streamflow forecasts utilizing multiple climate models and hydrologic models. *Water Resources Research*.-50(2): 1288–1307.
- See L, Openshaw S. 2000. A hybrid multi-model approach to river level forecasting. *Hydrological Science Journal*. 45(4):523–536.
- Shamseldin A, O'Connor K, Liang G. 1997. Methods for combining the output of different rainfall-runoff models. *Journal of Hydrology*. 197: 203–229.
- Tan B O, O'Connor KM. 1996. Application of an empirical infiltration equation in the SMAR

- conceptual model. *Journal of Hydrology*. 185: 275–295.
- Traore VB, Sambou S, Tamba S, Fall S, Diaw A T, Cisse MT. 2014. Calibrating the rainfall-runoff model GR4J and GR2M on the Koulountou river basin, a tributary of the Gambia River. *American Journal of Environmental Protection*. 3(1): 36–44.
- Tuteja NK, Cunnane C. 1999. A quasi physical snowmelt run-off modelling system for small catchments. *Hydrological Processes*. 13(12/13): 1961–1975.
- Vaze J, Chiew FHS, Perraud JM, Viney N, Post D, Teng J, Wang B, Goswami M. 2010. Rainfall-runoff modeling across southeast Australia: Datasets, models and results, *Australian Journal of Water Resources*. 14(2): 101–116.
- Zhang X, Waters D, Ellis R. 2013. Evaluation of SimHyd, Sacramento and GR4J rainfall runoff models in two contrasting Great Barrier Reef catchments. 20th International Congress on Modelling and Simulation, Adelaide, Australia. 3260–3266.
- Zhang L, Yang X. 2018. Applying a multi-model ensemble method for long-term Runoff prediction under climate change scenarios for the Yellow River Basin, China. *Water*. 10 (3).



Watershed Management Research

VOL. 32, No. 3, Ser. No: 124, Autumn 2019, pp. 63 -77
DOI: 10.22092/wmej.2019.124804.1187

Evaluating the Efficiency of K nearest Neighbor and Fuzzy C-Means Clustering Based Methods in the Outputs of Hydrological Models

Rozhin Delpasand

Watershed and Range Management Department, Gonbad Kavous University, Gonbad Kavous, Golestan

Aboalhasan Fathabadi*

(Corresponding Author)* Watershed and Range Management Department, Gonbad Kavous University, Gonbad Kavous, Golestan

Hamed Rouhani

Watershed and Range Management Department, Gonbad Kavous University, Gonbad Kavous, Golestan

Seyed Morteza Seyedian

Watershed and Range Management Department, Gonbad Kavous University, Gonbad Kavous, Golestan

*Corresponding Author Email: ahfathabadi@gmail.com

Received: 03 January 2019

Accepted: 18 June 2019

Abstract

Because of incomplete model input and imperfections of the model structure there is no single hydrological model that has the best performance in different conditions and present outputs without uncertainty. In this situation by combining individual model's outputs, the strengths of each single model are used to make a new model that performs better than each single model. The efficiency of nonparametric K nearest neighbor and the fuzzy C-means clustering based methods were compared with BGA (Bates Granger Averaging), GRA (Granger Ramanathan Averaging), AICA (Akaike Information Criterion), BICA (Bayes Information Criterion), equal weights averaging and lasso methods in averaging output of hydrological models GR5J, SimHyd, SACRAMENTO and SMAR. Firstly, using the amount of rainfall, evapotranspiration, temperature, and the daily discharge of the Kasilian Watershed in Pol Sefid city at the Bon Koh Station was simulated by each hydrological model. Then different model averaging methods were used to combine the output of each single model. Results indicated that for the calibration period, the GR5J and SACRAMENTO, and the correlation coefficient, Nash Sutcliffe efficiency and RMSE were 0.83, 0.69 and 0.24, respectively. The SimHyd and GR5J models performed better for the validation period; the correlation coefficient, Nash Sutcliffe efficiency and RMSE were 0.73, 0.27 and 0.52, respectively. The lasso and GRA model averaging had the best performance for the calibration period, and for the validation data equal weights averaging and BGA had the best performance. For calibration data K nearest neighbor performed better than fuzzy K means clustering based method and the best performance for two methods was obtained at 20 neighbors and for validation data fuzzy K means clustering based method performed better and its observed model performance was improved as the number of neighbors was increased. .

■ **Keywords: GR5J model, Kasilian, model combining, rainfall-runoff, SACRAMENTO model** ■