نشریه علمی پژوهشی مهندسی آبیاری و آب ایران
سال دهم. شماره سی و هفتم پاییز ۱۳۹۸

توسعه مدل سری زمانی ARIMA فصلی و بررسی عملکرد مدل در شبیه-سازی بارندگی ماهانه

لاله پرویز[۱]

تاریخ ارسال:۱۳۹۶/۰۱۲/۱۵
تاریخ پذیرش:۱۳۹۷/۰۷/۰۸

مقاله برگرفته از پژوهش مستقل از پایان‌نامه و طرح پژوهشی
پژوهشی

چکیده

بارندگی یکی از مولفه‌های مهم چرخه هیدرولوژیکی است که فرآیندهای سطحی و اتمسفریک را بهم مرتبط می-سازد. بنابراین مدلسازی و برآورد دقیق پارامتر در مدیریت منابع آب، برنامه‌ریزی آبیاری، مدیریت کشاورزی و تخصیص آب مورد نیاز است. مدل SARIMA از مدل‌های مرسوم در شبیه‌سازی بارندگی ماهانه می‌باشد. از نکات ضعف مدل نادیده گرفتن تغییرات بین ماه در هر سال است. بنابراین هدف این مقاله، توسعه مدل SARIMA با در نظر گرفتن تغییرات بین‌سالی، بین‌ماه و مقایسه عملکرد آن با مدل SARIMA در ایستگاه اردبیل می‌باشد. تحلیل خوشه‌ای با روش Ward's جهت خوشه‌بندی سری‌های زمانی بارندگی ماهانه و مدل رگرسیون خطی جهت تعیین رابطه بین مشخصه‌های آماری هر خوشه و مقادیر بارندگی ماهانه استفاده شدند. درصد کاهش آماره‌های RRMSE، RMSE، MAE از مدل SARIMA به مدل SARIMA توسعه‌یافته به‌ترتیب ۲۴/۰۵، ۱۷/۲۴ و ۲۸/۴۸، بیانگر عملکرد قابل قبول مدل توسعه‌یافته بود. در مقایسه مقادیر مشاهداتی و شبیه‌سازی، مدل توسعه‌یافته دارای تخمین بیش‌برآورد می-باشد. کاهش ۵۱/۱۶ درصد RRMSE در شبیه‌سازی بارندگی فصلی نیز بیانگر افزایش دقت مدل توسعه‌یافته براساس تحلیل خوشه‌ای بود. ضریب همبستگی بین بارندگی شبیه‌سازی با مدل SARIMA توسعه‌یافته و مشاهداتی به مقدار زیادی افزایش یافته و به سطح معنی‌داری رسیده است. بنابراین استفاده از تحلیل خوشه‌ای و توسعه مدل SARIMA میزان دقت عملکرد مدل را افزایش داده است.

واژه‌های کلیدی: بارندگی، توسعه، تحلیل خوشه‌ای، SARIMA.

[۱] لاله پرویز (استادیار مهندسی منابع آب)، دانشکده کشاورزی، دانشگاه شهید مدنی آذربایجان، تبریز، ایران.
تلفن تماس نویسنده (۰۹۱۴۴۱۴۶۲۴۶). آدرس پست الکترونیکی نویسنده (laleh_parviz@yahoo.com) (نویسنده مسئول)

مقدمه

بارندگی یکی از مولفه‌های پیچیده و مهم چرخه هیدرولوژیکی می‌باشد (Aman, 2017). بنابراین برآورد دقیق آن در برنامه‌ریزی و مدیریت منابع آب از اهمیت چشمگیری برخوردار است. روش‌های آماری قلمرو و دقت شبیه‌سازی مدل را افزایش می‌دهند. مدلسازی آماری و تحلیل داده از ابزار کلیدی در مطالعه فرآیندهای استوکستیکی است. مدل میانگین متحرک تجمعی خودبازگشت (ARIMA[1]) به‌طور وسیعی در این زمینه کاربرد دارد. مدل ARIMA یک نوع از مدل‌های پیش‌بینی کوتاه مدت در تحلیل سری‌های زمانی است و روش نسبتا سیستماتیک و انعطاف‌پذیری می‌باشد. به منظور در نظر گرفتن تغییرات فصلی، فرم کلی مدل ARIMA به مدل ARIMA فصلی (SARIMA[2]) تغییر یافت (Narasimha Murthy et al., 2018). یکی از نکات ضعف مدل SARIMA، مربوط به تغییرات زمانی است که جهت رفع این مشکل مدل ISARIMA[3] پیشنهاد شد که هدف افزایش میزان دقت در مقادیر شبیه‌سازی می‌باشد.

حسینعلی‌زاده و همکاران (۲۰۱۴) از آمار ۳۰ ساله بارندگی ماهانه در ایستگاه ارازکوسه استان گلستان به‌عنوان یک سری زمانی جهت مدلسازی استفاده کردند. براساس نمودارهای خودهمبستگی[4] و خودهمبستگی جزیی[5]، الگوی مناسب به‌صورت $SARIMA(0,0,1)\times(0,1,1)_{12}$ به داده‌ها برازش داده شد. نتایج تحقیق حاکی از مناسب بودن الگوی SARIMA در پیش‌بینی بارندگی ماهانه استان است. عبدالله نژاد (۸) به پیش‌بینی بارندگی ماهانه در ایستگاه هاشم آباد گرگان در طی دوره آماری ۲۰۱۳-۲۰۱۵ پرداختند. بعد از بررسی باقی‌مانده‌ها و خطاهای پیش‌بینی، مدل منتخب شامل $SARIMA(3,1,2)\times(0,1,2)_{12}$ بود. نتایج بیانگر این است که مدل SARIMA توانسته روند تغییرات سری زمانی را با خطای کمتری شبیه‌سازی کند. در هر دو تحقیق تاکید بر ارائه روند مدلسازی با مدل SARIMA بود و مقایسه جامعی بین آماره‌های خطا انجام نگرفته است. Dabral and Murry (2017) از مدل SARIMA جهت مدلسازی و پیش‌بینی بارندگی در منطقه‌ای در شمال هند استفاده کردند. میانگین و انحراف معیار مقادیر پیش‌بینی و مشاهداتی اختلاف کمی داشتند. مقادیر ضریب نش[6] نیز حاکی از برازش مدل با دقت بالا به داده‌های مشاهداتی بود. Narasimha Murthy et al. (2018) به مدلسازی و پیش‌بینی الگوی بارش در هند با مدل SARIMA پرداختند. بعد از طی مراحل مدلسازی باکس- جنکینز، مدل منتخب، مدل ARIMA $(0,1,1)\times(1,0,2)4$ بود. برازش مدل به داده‌های مشاهداتی خوب بود و مدلسازی نوسانات فصلی بجز در برخی موارد حدی موفق بوده است. نتایج شبیه‌سازی نشان داد الگوی بارندگی در سه سال صحت‌سنجی یکسان می‌باشد. Wang et al. (2014) به مقایسه عملکرد مدل ARIMA توسعه‌یافته با مدل‌های مرسوم سری زمانی در شبیه‌سازی بارندگی ایستگاه Lanzhou در چین پرداختند. طول دوره آماری مورد استفاده مربوط به واسنجی مدل، ۱۹۵۱ – ۲۰۰۰ و صحت‌سنجی ۲۰۰۱ بود. مدل‌های مورد استفاده جهت مقایسه با عملکرد مدل ARIMA توسعه یافته شامل مدل SARIMA، مدل ARIMA برای هر ماه بطور جداگانه، AR(24) و AR(36) بود. نتایج نشان داد دقت مدل ARIMA توسعه‌یافته به‌طور معنی‌داری نسبت به مدل فصلی افزایش یافته است. کمترین میانگین خطای مطلق مربوط به مدل ARIMA توسعه‌یافته با مقدار ۹/۴۱ میلیمتر و ۲۱ درصد افزایش دقت پیش‌بینی بود.

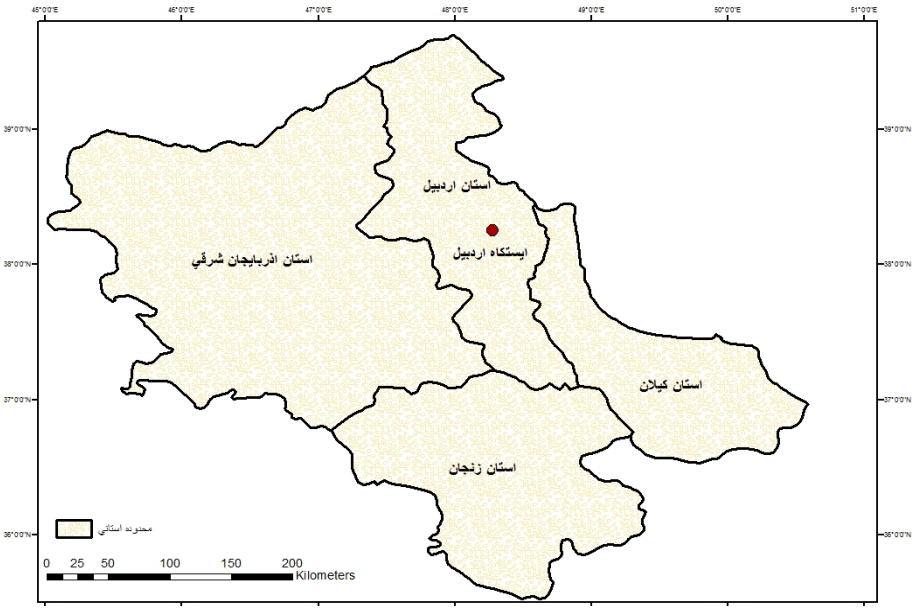
[1] Autoregressive Integrated Moving Average
[2] Seasonal Autoregressive Integrated Moving Average
[3] Improved Autoregressive Integrated Moving Average
[4] Autocorrelation Function (ACF)
[5] Partial Autocorrelation Function (PACF)
[6] Nash-Sutcliffe of efficiency

در بیشتر تحقیقات انجام گرفته، در مدلسازی فصلی سری زمانی از مدل SARIMA استفاده شده است بنابراین توسعه مدل سری زمانی فصلی بخصوص در کشور ایران که تحقیقی در زمینه توسعه مدل SARIMA انجام نگرفته است، جهت برنامه‌ریزی و مدیریت دقیق منابع آب ضروری بنظر می‌رسد. هدف این مقاله توسعه مدل SARIMA با در نظر گرفتن تغییرات زمانی (تغییرات بین‌سالی و بین‌ماه) با رویکرد خوشه‌بندی و استفاده از مدل رگرسیون خطی در مدلسازی بارندگی ماهانه و مشخصه‌های آماری بود. عملکرد مدل SARIMA توسعه‌یافته با استفاده از داده‌های بارندگی ایستگاه اردبیل مورد ارزیابی و با مدل SARIMA مورد مقایسه قرار گرفت.

مواد و روش ها

منطقه مورد مطالعه

داده‌های مورد استفاده تحقیق مربوط به بارندگی ایستگاه هواشناسی اردبیل در طی دوره آماری ۴۱ ساله است. موقعیت ایستگاه اردبیل در استان اردبیل و استان‌های مجاور در شکل ۱ نشان داده شده است. براساس محاسبه شاخص دومارتن، اقلیم منطقه در محدوده نیمه‌خشک قرار دارد. با توجه به تغییرات اقلیمی حاکم در منطقه (روند بارش با آزمون من-کندال محاسبه شده در تحقیق در منطقه منفی است)، موقعیت استراتژیکی منطقه و نقش استان در تولید محصولات کشاورزی در کشور، برآورد بارندگی با دقت بالا از اهمیت چشمگیری برخوردار است.



شکل (۱): موقعیت مکانی ایستگاه مطالعاتی

مدل SARIMA

مدل SARIMA از مدل‌های خطی عمومی برای پیش‌بینی سری زمانی فصلی است (Chen and Wang, 2007). شکل عمومی مدل به‌صورت SARIMA(p,d,q)×(P,D,Q)W است که بخش دوم مدل جهت در نظر گرفتن بخش فصلی سری زمانی می‌باشد. فرم کلی مدل برساس رابطه ۱ است.

$$\Phi_{\mathrm{P}}(B^{\omega})\,\phi_{p}(B)\nabla_{\omega}^{D}\,\nabla^{d}z_{\mathrm{t}} = \Theta_{\mathrm{Q}}(B^{\omega})\,\theta_{\mathrm{q}}(B)\,\varepsilon_{t} \qquad (۱)$$

p: مرتبه مدل خودهمبسته غیرفصلی، q: مرتبه مدل میانگین متحرک غیر فصلی، P: مرتبه مدل خودهمبسته فصلی، Q: مرتبه مدل میانگین متحرک فصلی، B: عملگر تفاضل، ∇^{d} : d امین تفاضل غیرفصلی، ∇_{ω}^{D}: D امین تفاضل فصلی به اندازه w، ϕ: پارامتر مدل خودهمبسته غیرفصلی، θ: پارامتر مدل میانگین متحرک غیرفصلی، Φ: پارامتر مدل خودهمبسته فصلی، Θ: پارامتر مدل میانگین متحرک فصلی (کارآموز و عراقی‌نژاد، ۱۳۸۵). روندنمای مدلسازی سری‌های زمانی براساس تحلیل اولیه سری زمانی، تعیین مرتبه‌های مدل، تخمین پارامتر و آزمون نکویی برازش مدل انجام می‌گیرد (Carmen Bass et al., 2017).

مدل SARIMA توسعه‌یافته

مدل SARIMA برای سری‌های زمانی با فواصل یا گام‌های زمانی مشخصی مانند ماه کاربرد دارد. از آنجایی که مدل SARIMA توانایی ارتباط با تغییرات بین‌سالی هر ماه را در سری زمانی ماهانه دارد، اطلاعات مربوط به تغییرات بین ماه سری زمانی ممکن است از بین برود. به‌عنوان نمونه بعد از تفاضل فصلی از مرتبه ۱۲ یک سری زمانی ماهانه، سری‌های زمانی اصلی به یک سری زمانی جدید بدون دوره فصلی تبدیل می‌شود. مدل ARIMA غیر فصلی به سری زمانی جدید برازش داده می‌شود به طوری که تغییرات بین‌ماه سری‌های زمانی ماهانه اصلی به تغییرات بین‌ماه سری جدید بعد از تفاضل‌گیری تبدیل شده‌اند. تبدیل تغییرات بین‌ماه سری زمانی ماهانه اصلی به تغییرات بین‌ماه جدید حاصل از تفاضل‌گیری فصلی ممکن است منجر به از دست دادن دقت مدل شود. به‌منظور جلوگیری از دست دادن اطلاعات تغییرات بین‌ماه، مدل SARIMA جهت در نظر گرفتن همزمان هر دو نوع تغییر زمانی (تغییرات بین‌سالی و بین‌ماه)، توسعه داده شد. تحلیل خوشه‌ای جهت طبقه‌بندی سری‌های زمانی ماهانه و استخراج مشخصه‌های آماری هر کلاس داده‌ها استفاده شد. مشخصه‌های هر کلاس شامل بیشینه، کمینه و میانگین پیراسته سری‌های هر خوشه بودند. سپس مدل رگرسیون خطی بین سرهای زمانی ماهانه و مشخصه‌های آماری هر خوشه ساخته شد. در این مدل رگرسیونی داده‌های ماهانه بارندگی به‌عنوان متغیر وابسته و مشخصه‌های آماری به عنوان متغیر مستقل در نظر گرفته شدند. سپس از مدل ARIMA جهت مدلسازی مشخصه‌های آماری هر کلاس استفاده شد که تغییرات بین‌ماه هر مشخصه آماری در نظر گرفته می‌شود. این روش سعی دارد مانع از بین رفتن اطلاعات مربوط به تغییرات بین‌ماه شود. پیاده‌سازی مدل ARIMA توسعه‌یافته دارای چندین مرحله است که مراحل در شکل ۲ نشان داده شده است. براساس شکل ۲ مراحل مدلسازی مدل SARIMA توسعه‌یافته بصورت زیر است.

i. انجام تحلیل خوشه‌ای بر داده‌های ماهانه و گروه‌بندی ماه‌ها با تغییرات هیدرولوژیکی مشابه

ii. پیدا کردن مشخصه‌های آماری (بیشینه، کمینه و میانگین پیراسته) هر خوشه

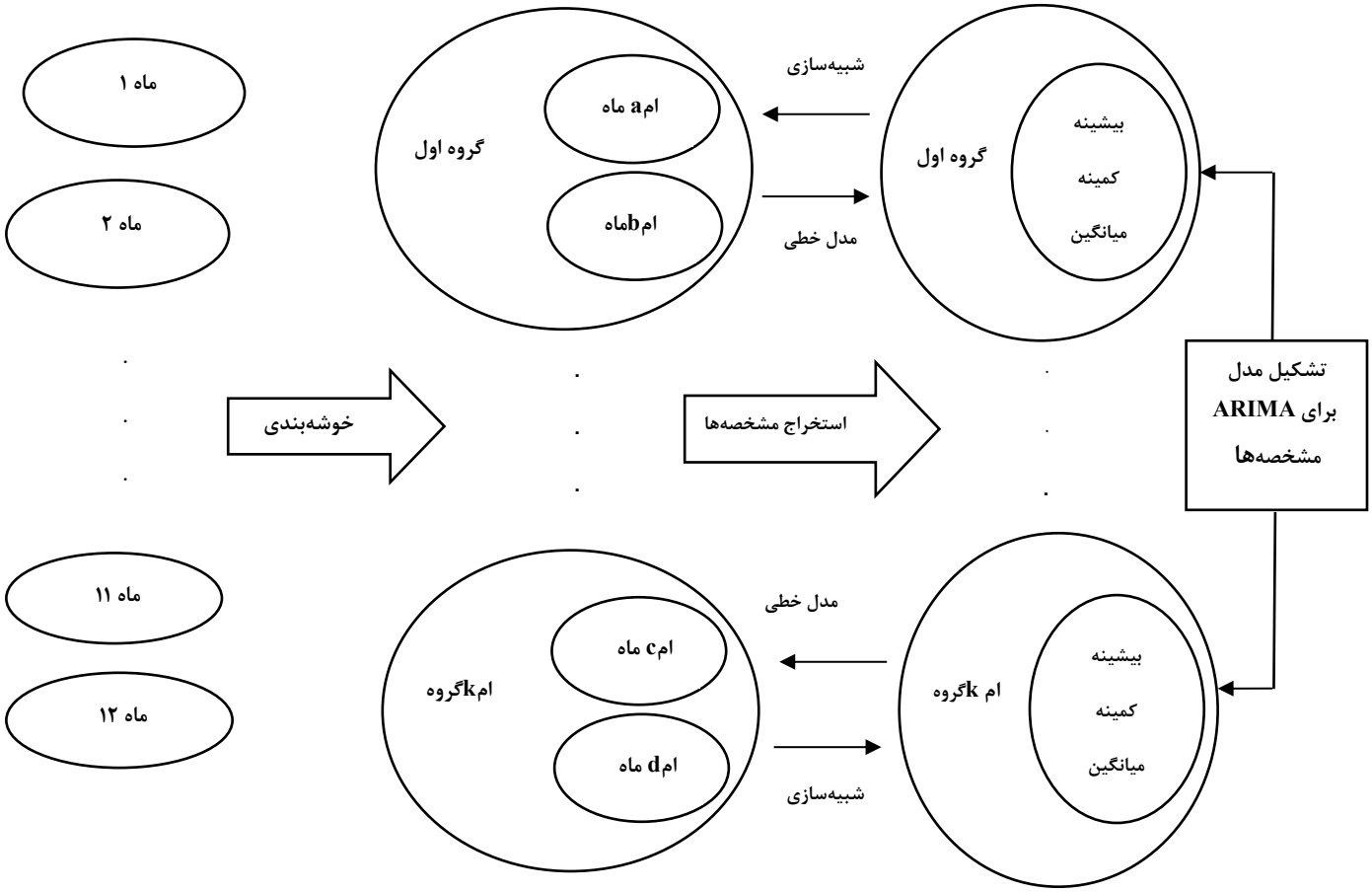
iii. ساخت مدل رگرسیون خطی، به‌عنوان نمونه بارندگی در ماه i ام از رابطه ۲ پیروی می‌کند.

$$y_i = a_i\,y_{j,\max} + b_i\,y_{j,\min} + c_i\,y_{j,avg} + d_i \qquad (۲)$$

a_i، b_i، c_i و d_i: ضرایب مدل برای بارندگی ماه i ام، $y_{j,max}$، $y_{j,min}$ و $y_{j,avg}$ به ترتیب بیشینه، کمینه و میانگین پیراسته کلاس j ام.

iv. ساخت مدل ARIMA برای مشخصه‌های آماری (بیشینه، کمینه و میانگین پیراسته) هرخوشه و پیش‌بینی مشخصه‌های آماری برای دوره صحت‌سنجی

v. جایگزینی مشخصه‌های پیش‌بینی شده در مدل رگرسیون خطی ساخته شده در معادله ۲ و بدست آوردن سری‌های زمانی بارندگی هر ماه (Wang et al., 2014).



شکل (۲): مراحل شبیه‌سازی مدل SARIMA توسعه یافته براساس تحلیل خوشه‌ای و رگرسیون خطی

ارزیابی عملکرد مدل ARIMA توسعه‌یافته

جهت مقایسه عملکرد مدل ARIMA توسعه‌یافته با مدل ARIMA فصلی از برخی آماره‌ها استفاده شده است. آماره‌های مورد استفاده شامل خطای ریشه متوسط مربعات (RMSE[1]) که مقایسه بین مقادیر مشاهداتی و شبیه‌سازی است، خطای ریشه متوسط مربعات نسبی (RRMSE[2]) که در واقع این معیار، بی-بعد شده معیار RMSE است که از تقسیم معیار RMSE بر میانگین داده‌های مشاهداتی بدست می‌آید و میانگین خطای مطلق (MAE[3]) در واقع بی‌بعد شده خطای بین مقادیر مشاهداتی و شبیه‌سازی است.

[1] Root Mean Square Error

[2] Relative Root Mean Square Error

[3] Mean Absolute Error

شاخص سازش[1] معیار دیگری برای ارزیابی عملکرد مدل است که فرم ریاضی آن در رابطه ۳ نشان داده شده است.

$$d = 1 - \frac{\sum_{i=1}^{n}(O_i - S_i)^2}{\sum_{i=1}^{n}(|S_i - \overline{O}| + |O_i - \overline{O}|)^2} \qquad (۳)$$

O_i: مقادیر مشاهداتی، S_i: مقادیر شبیه‌سازی، n: تعداد داده‌ها، $\overline{O}$: میانگین مقادیر مشاهداتی.

شاخص سازش می‌تواند بر مشکل عدم حساسیت ضریب نش و ضریب تعیین نسبت به تفاوت در میانگین و واریانس مشاهده شده غلبه کند. دامنه شاخص بین صفر تا یک است و مقادیر نزدیک یک بیانگر عملکرد دقیق مدل شبیه‌سازی می‌باشند (Willmott et al., 2011).

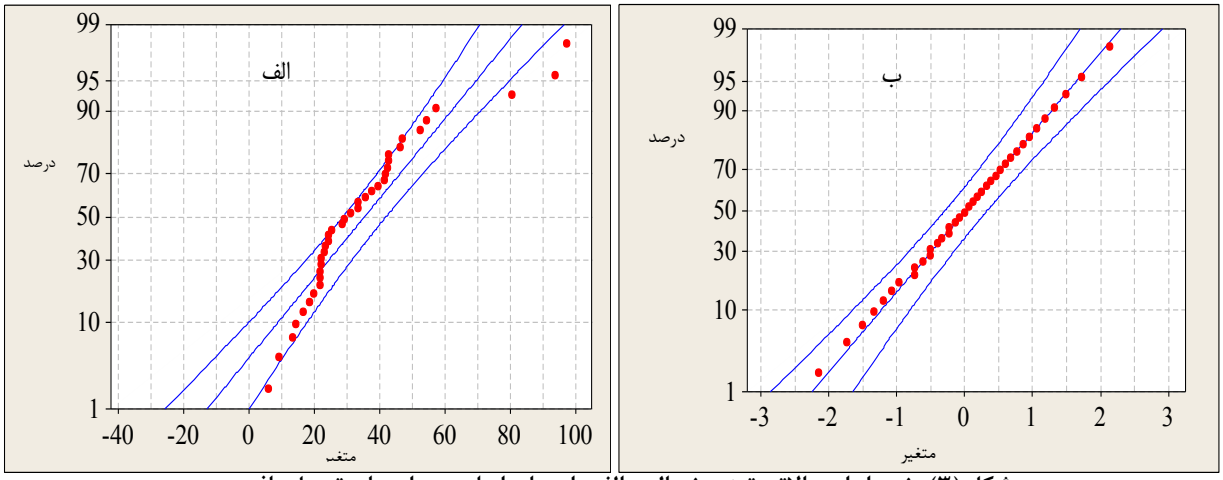
نتایج

سری زمانی مورد استفاده شامل سری بارندگی ماهانه طی دوره آماری ۱۹۷۶-۲۰۱۳ بود که دوره صحت‌سنجی مربوط به ۲۰۱۴-۲۰۱۶ بود. براساس پایه تغییرات فصلی داده‌های بارندگی از مدل SARIMA استفاده شد و سپس مدل براساس تحلیل خوشه‌ای توسعه داده شد. در ادامه نتایج حاصل از شبیه‌سازی بارندگی با مدل SARIMA و SARIMA توسعه‌یافته بیان می‌شود.

مدل ARIMA فصلی

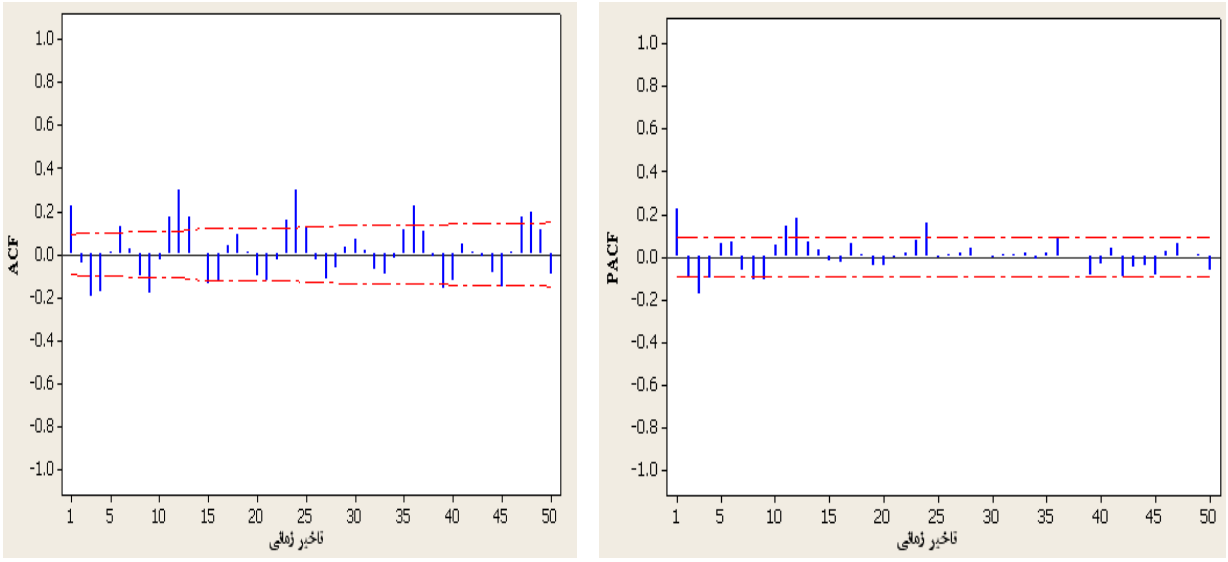
روند مدلسازی استوکستیکی شامل پیروی از مراحل مربوط به روند نمای باکس- جنکینز است. بررسی مستقل و نرمال بودن سری‌های زمانی از مراحل اولیه مدلسازی می‌باشد. بررسی نرمال بودن سری زمانی ماهانه حاکی از نیاز به نرمال‌سازی سری‌های زمانی ماهانه بود. به‌عنوان نمونه سری زمانی بارندگی ماه March قبل و بعد از نرمال‌سازی در شکل ۳ آورده شده است.

[1] Index of Agreement (d)



شکل (۳): نمودار احتمالاتی توزیع نرمال – الف داده‌های اصلی ب داده‌های تبدیل یافته

مقایسه دو شکل ۳ الف و ب حاکی از آن است که با استفاده از تبدیل مناسب داده‌ها و نرمالسازی، داده‌ها در محدوده‌های اطمینان قرار می‌گیرند. در مرحله بعد سری زمانی باید از نظر تصادفی بودن مورد بررسی قرار گیرد. با آزمون استقلال در زمان و استفاده از تابع خودهمبستگی سری زمانی، تصادفی بودن سری زمانی مورد ارزیابی قرار گرفت. در شکل ۴ توابع خودهمبستگی و خودهمبستگی جزیی با محدوده اطمینان ۹۵ ٪ سری زمانی ترسیم شده است.

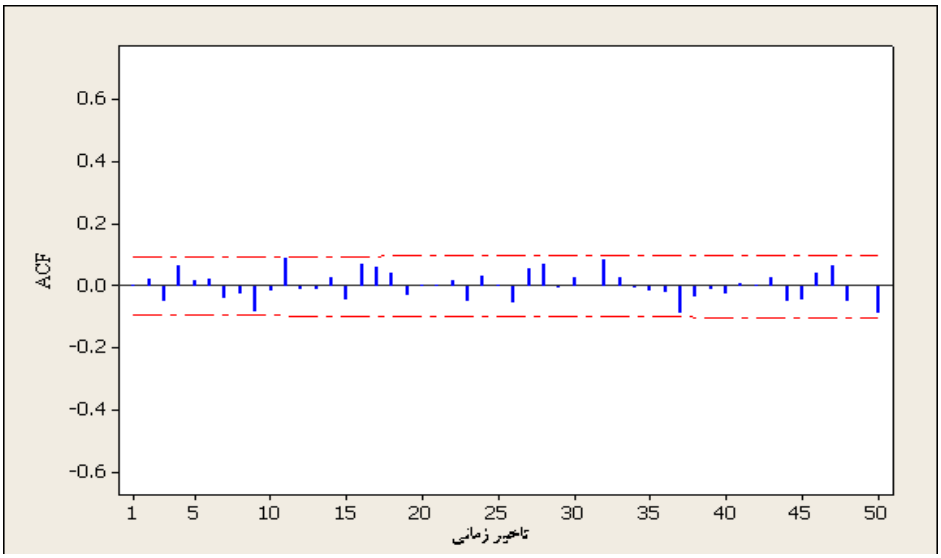


شکل (۴): توابع خودهمبستگی و خودهمبستگی جزیی سری زمانی

آنچه از ترسیم توابع خودهمبستگی قابل حصول است مربوط به دوره تناوب سری‌های زمانی است که دوره تناوب سری‌های ترسیم شده ۱۲ است. در بررسی دیگر به دلیل قرارگیری مقادیر توابع در تاخیرهای مختلف در خارج از محدوده اطمینان وابستگی زمانی در سری‌ها کاملا آشکار است. در مرحله بعد باید روند

داده‌ها مورد بررسی قرار گیرد که از آزمون های مورد استفاده در این قسمت، آزمون من-کندال فصلی است (بهمنی و همکاران، ۱۳۹۲). نتایج حاصل از آزمون من-کندال فصلی حاکی از عدم وجود روند معنی‌دار در سری‌های زمانی ماهانه بود، مقدار آماره Z از مقدار بحرانی آن کمتر بود. بنابراین تفاضل‌گیری غیرفصلی لازم نمی‌باشد (d=0) ولی به علت دوره‌ای بودن سری-های زمانی از تفاضل‌گیری فصلی سری زمانی استفاده شد. مساله قابل توجه در نتایج بررسی روند حاکی از آن است که با وجود عدم روند معنی‌داری داده‌ها، علامت روند سری زمانی بارندگی منفی است که این مساله می‌تواند ناشی از تغییرات اقلیمی باشد. مرحله بعدی مدلسازی سری زمانی مربوط به شناسایی مدل است. معیار آکائیک تعداد پارامترهای مناسب خودبازگشت را بیشتر از حد واقعی برآورد می‌کند، بنابراین در تعیین مرتبه‌های مدل از معیار SBC[1] استفاده شد (مالمیر، ۱۳۸۵). برای این منظور معیار SBC برای مرتبه‌های مدل با دامنه صفر تا سه (برای مرتبه خودهمبسته غیرفصلی: ۰-۳، میانگین متحرک غیر فصلی: ۰-۳، خودهمبسته فصلی: ۰-۳ میانگین متحرک فصلی ۰-۳) محاسبه شدند. مرحله بعد شامل بررسی آزمون نکویی برازش مدل‌ها با مرتبه‌های مختلف (مدل‌هایی که مقادیر SBC آنها محاسبه شده است) می‌باشد. از بین مدل‌ها با مرتبه‌های گوناگون که آزمون نکویی برازش را گذرانده‌اند، مدل منتخب مربوط به مدلی با کمینه مقدار SBC است. مدل منتخب سری زمانی بارندگی اردبیل شامل مدل $ARIMA(3,0,2)\times(0,1,2)_{12}$ با کمینه مقدار SBC برابر با ۲/۷۷- است. آزمون نکویی برازش مدل منتخب که شامل بررسی نرمال و مستقل بودن باقی‌مانده‌های مدل است در ادامه توضیح داده می‌شود. سطح معنی‌داری مربوط به بررسی نرمال بودن باقی‌مانده‌های مدل با آزمون کولموگروف-اسمیرنوف و Shapiro-Wilk انجام گرفت. سطح معنی‌داری در آزمون کولموگروف- اسمیرنوف و Shapiro-Wilk به ترتیب برابر با ۰/۲ و ۰/۱۹۸ بوند که در هر دو آزمون مقادیر از مقدار ۰/۰۵ بزرگتر می‌باشند و بیانگر نرمال بودن داده‌ها هستند. بررسی استقلال سری زمانی باقی مانده‌های مدل با رسم تابع خودهمبستگی با محدوده اطمینان ۹۵ ٪ در شکل ۵ آورده شده است.

[1]- Schwarz Bayesian Criteria

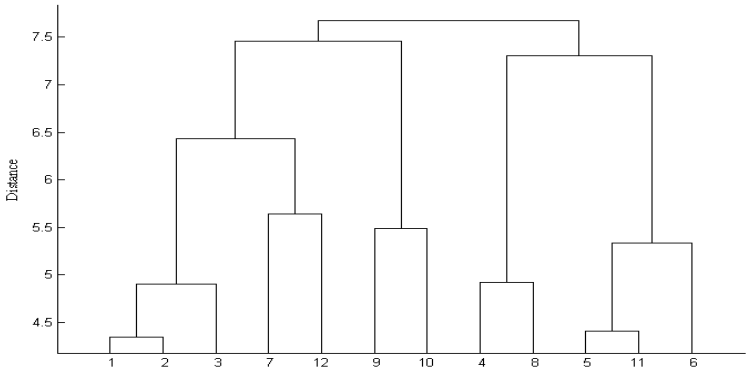


شکل (۵): تابع خودهمبستگی سری زمانی باقی‌مانده‌های مدل منتخب

با توجه به قرار گیری تابع خودهمبستگی در محدوده‌های اطمینان، مستقل بودن سری زمانی باقی مانده‌های مدل منتخب تایید می‌شود. بعد از تعیین مدل منتخب، اقدام به شبیه‌سازی در دوره صحت-سنجی می‌شود که نتایج عملکرد در بخش‌های بعدی آورده شده است.

مدل SARIMA توسعه‌یافته

مرحله اول مدلسازی با مدل SARIMA توسعه-یافته مربوط به خوشه‌بندی سری زمانی ماهانه بارندگی است یعنی قرار گیری سری‌های زمانی ماهانه با خصوصیات مشابه هم در یک دسته. در این تحقیق تحلیل خوشه‌ای با فاصله اقلیدسی و روش Ward's انجام گرفت که نمودار درختی سری‌های زمانی ماهانه بارندگی در شکل ۶ آورده شده است.



شکل (۶): درخت خوشه‌بندی سری‌های زمانی بارندگی

نتایج

به ۶

این‌صورت است: خوشه ۱: January-February-March-July-December خوشه ۲: September-October خوشه ۳: April-August خوشه ۴: May-June-November. قابل ذکر است که برخی از دیگر روش‌ها مانند نزدیک‌ترین و دورترین همسایگی نیز مورد بررسی قرار گرفت که خوشه‌بندی روش Ward's از انسجام کامل‌تری برخوردار بود. در مرحله بعد اقدام به ساخت مدل رگرسیون برای ایجاد ارتباط بین مشخصه‌های آماری (بیشینه، کمینه، میانگین پیراسته) و سری‌های زمانی هر ماه در هر خوشه شد که نتایج پارامترهای تخمینی در جدول ۱ آورده شده است.

جدول (۱): پارامترهای تخمینی مدل رگرسیون خطی

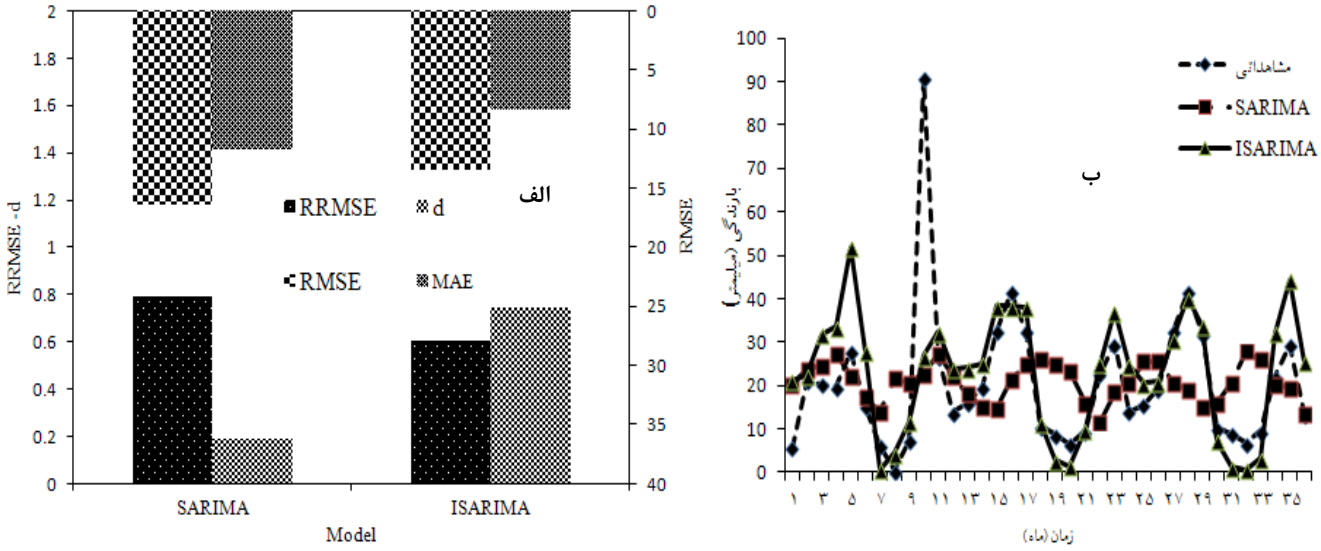
خوشه	ماه	a_i	b_i	c_i	d_i
خوشه ۱	January	۰/۰۲۳	-۰/۲۱۱	۰/۸۲	۵/۲۹
	February	-۰/۲۲۸	-۱/۲	۱/۴۶	۵/۲۴
	March	۰/۳۴	-۰/۵۲	۱/۳۷	-۷/۰۳
	July	-۰/۱۱	۱/۷۷	۰/۱۶	۱/۷۸
	December	۰/۸۲۲	۰/۸۷۲	-۱/۴	۱۵/۳۷
خوشه ۲	September	۰	۱/۱۲	-۰/۰۴۴	۲/۴۹
	October	۰	-۱/۱۲	۲/۰۴	-۲/۴۹
خوشه ۳	April	۰	-۱/۰۵	۲/۰۴	-۱/۱۵
	August	۰	۱/۰۵	-۰/۰۴۲	۱/۱۵
خوشه ۴	May	۱/۱۷	۱/۳۲	-۱/۶۹	۷/۶۲
	June	-۰/۱۱	۱/۱۳	۰/۱۲۸	۵/۹۴
	November	۰/۱۱	-۱/۱۳	۱/۸۷	-۵/۹۴

بعد از تعیین پارامترهای تخمینی در جدول ۱، معادلات رگرسیون خطی لازم در مرحله ۳ مدلسازی SARIMA توسعه‌یافته آماده می‌باشد. قابل ذکر است بررسی احتمال معنی‌داری معادلات رگرسیونی در سطح اطمینان ۵ درصد حاکی از برازش دقیق مدل رگرسیونی بر داده‌های مورد استفاده می‌باشد و در نهایت مدل‌های رگرسیونی برازشی مدل‌های مناسبی هستند. سپس جهت مدلسازی مشخصه‌های آماری هر خوشه، از مدل ARIMA استفاده شد که نتایج مدل-های منتخب در جدول ۲ آورده شده است.

جدول (۲): مرتبه مدل‌های برازشی بر مشخصه‌های آماری هر خوشه

مشخصه‌های آماری	خوشه ۱			خوشه ۲			خوشه ۳			خوشه ۴		
	بیشینه	کمینه	میانگین پیراسته	بیشینه	کمینه	میانگین پیراسته	بیشینه	کمینه	میانگین پیراسته	بیشینه	کمینه	میانگین پیراسته
p	۲	۱	۱	۲	۳	۱	۳	۱	۱	۱	۲	۲
q	۲	۳	۲	۲	۲	۰	۲	۳	۱	۳	۳	۲

لازم به ذکر است که بررسی وجود روند در مشخصه‌های آماری حاکی از عدم وجود روند در سری-های زمانی مشخصه‌های آماری است. بعد از تعیین مدل‌های منتخب، اقدام به برآورد مقادیر مشخصه‌های آماری در طی دوره صحت‌سنجی شد که بعد از جایگذاری در معادلات رگرسیونی بدست آماده در هر خوشه، مقادیر سری زمانی بارندگی در هر ماه بدست آمد. نتایج مقایسه آماره‌های بررسی عملکرد مدل SARIMA و SARIMA توسعه‌یافته در شکل ۷ آورده شده است .



شکل (۷): مقایسه عملکرد دو مدل (الف) نمودار بارندگی مشاهداتی و شبیه سازی (ب)

SARIMA توسعه‌یافته کاملا مشهود است به-طوری که درصد کاهش آماره‌های RRMSE، RMSE، MAE از مدل SARIMA به مدل SARIMA توسعه‌یافته به‌ترتیب ۲۴/۰۵، ۱۷/۲۴ و ۲۸/۴۸ بوده است و این مساله تاثیر خوشه‌بندی و قرار گرفتن داده-های هیدرولوژیکی با مشخصات یکسان را در خوشه‌ها نشان می‌دهد. Wang et al. (2014) نیز با استفاده از تحلیل خوشه‌ای و اعمال رگرسیون خطی در مدل SARIMA به کمترین میزان خطا در مدل دست یافتند. در شکل ۷-ب نیز در بیشتر ماه‌های مربوط به دوره صحت‌سنجی، نمودار بارندگی مشاهداتی و شبیه-سازی با مدل SARIMA توسعه‌یافته از تطابق بیشتری برخوردار است. همچنین متوسط بارندگی در طی دوره صحت‌سنجی بیانگر آن است که مدل توسعه‌یافته دارای بیش‌برآوردی در حدود ۱۰/۴ درصد است. مقایسه مقادیر RRMSE در بین ماه‌های دوره صحت‌سنجی حاکی از کمینه مقدار خطا به ترتیب در ماه‌های February (۰/۱۷)، April (۰/۲۴)، March (۰/۲۶) بود. تطابق زمانی ماه‌هایی با کمینه خطا بین تحقیق حاضر و تحقیق Wang et al. (2014) نیز مشاهده شد. همچنین در تحقیق Wang et al. (2014) بیان شد که ماه September دارای تخمین کم برآورد است، در این تحقیق نیز متوسط بارندگی در طی دوره صحت‌سنجی در ماه September دارای تخمین کم برآوردی با مقدار ۶/۳ درصد است. که این مساله تاثیر خوشه‌بندی در استخراج مشخصه‌های مشابه در داده ها و سری زمانی را نشان می‌دهد. در طی دوره صحت-سنجی، ضریب همبستگی بین مقادیر بارندگی

مشاهداتی و شبیه سازی با مدل SARIMA بسیار کم و غیرمعنی‌دار می‌باشد در حالی که ضریب همبستگی بین مقادیر بارندگی مشاهداتی و شبیه‌سازی با مدل SARIMA توسعه‌افته برابر با ۰/۸۱۱ و معنی‌دار در سطح یک درصد می‌باشد. در مقایسه دیگر عملکرد مدل به‌صورت فصلی مورد بررسی قرار گرفت به طوری- که میانگین بارندگی فصول پاییز، زمستان، بهار و تابستان با دو مدل مورد بررسی قرار گرفت. مقادیر RRMSE در مدل SARIMA برابر با ۰/۴۳ و در مدل SARIMA توسعه‌یافته برابر با ۰/۲۱که حاکی از کاهش ۵۱/۱۶ درصد خطا با توسعه مدل و مقادیر RMSE در مدل SARIMA برابر با ۹ و در مدل SARIMA توسعه‌یافته برابر با ۴/۳۹ بود که حاکی از کاهش ۵۱/۲۲ درصد خطا با توسعه مدل بود.

نتیجه‌گیری

توسعه مدل‌ها در جهت افزایش دقت شبیه‌سازی متغیرهای هیدرولوژیکی از مسایل پیشروی مهندسی منابع آب می‌باشد. یکی از روش‌های پیشرو تحلیل سری زمانی با روندنمای باکس- جنکینز و مدل ARIMA است. علیرغم کاربرد وسیع مدل ARIMA، در برخی از جهات جهت افزایش دقت شبیه‌سازی نیاز به توسعه مدل آشکار است. مدل SARIMA فصلی در پیش‌بینی سری‌های زمانی فصلی در بخش تحلیل زمانی دارای محدودیت می‌باشد. بنابراین جهت در نظر گرفتن تغییرات بین‌ماه و بین‌سالی، مدل SARIMA فصلی با استفاده از تحلیل خوشه‌ای توسعه داده شد. در تحلیل خوشه‌ای با تعیین گروه‌هایی از داده‌ها که بهم شبیه می‌باشند می‌توان رفتار دورنی سری‌ها را بهتر شناسایی کرد و بر مبنای خوشه‌بندی انجام شده مسیر فرآیند را تعیین کرد به طوری که بتوان نتایج بهتری را بدست آورد. تحلیل خوشه‌ای سلسله مراتبی مورد استفاده تحقیق نیز به خوشه‌های نهایی براساس میزان عمومیت آنها ساختار سلسه مراتبی، به‌صورت درختی نسبت می‌دهد (آرمان و همکاران، ۱۳۹۲). مقایسه آماره‌های ارزیابی عملکرد مدل و ترسیم نمودار بارندگی مشاهداتی و شبیه‌سازی بیانگر آن است که با توسعه مدل SARIMA میزان خطا کاهش و مقدار شاخص سازش افزایش یافته است. همچنین مقادیر شبیه- سازی‌شده ماهانه و فصلی مدل SARIMA توسعه‌یافته نسبت به مدل SARIMA از اختلاف کمتری نسبت به مقادیر مشاهداتی برخوردار بود. مقادیر ضریب همبستگی بین مقادیر مشاهداتی و شبیه‌سازی با مدل SARIMA توسعه‌یافته به مقدار زیادی افزایش یافته است و به سطح معنی‌داری نیز نزدیک شده است. در روند مدلسازی مقادیر روند حاصل از آزمون من- کندال منفی محاسبه شده است که می‌تواند ناشی از تغییرات اقلیمی حاکم باشد. براساس نتایج بدست آمده جهت افزایش دقت شبیه‌سازی مدل ISARIMA، برخی پیشنهادات وجود دارد از جمله استفاده از سایر مدل- های هوش مصنوعی مانند شبکه عصبی مصنوعی در مدلسازی بین مشخصه‌های آماری بین هر خوشه با سری‌های زمانی ماهانه، استفاده از سایر الگوریتم‌های خوشه‌بندی. با توجه به مقادیر آماره‌ها، عملکرد مدل SARIMA توسعه‌یافته گزینه مناسبی برای شبیه- سازی سری‌های زمانی بارندگی می‌تواند باشد.

منابع

آرمان، ن.، ع. سلاجقه، س. فیض نیا، ح. احمدی، ج. قدوسی و ع. کیانی‌راد. ۱۳۹۲. تعیین حوزه‌های آبخیز همگن جهت برآورد فرسایش و رسوب با استفاده از روش‌های مختلف تحلیل خوشه‌ای (مطالعه موردی: دامنه شمالی البرز). مرتع و آبخیزداری، مجله منابع طبیعی ایران. جلد ۶۹، شماره ۲، ص ۲۷۳-۲۶۱.

بهمنی، ر.، ف. رادمنش، س. اسلامیان و غ. پرهام. ۱۳۹۲. تحلیل روند تبخیر از مخزن سد و پیش بینی آن به کمک سری‌های زمانی. مجله علوم و مهندسی آبیاری. جلد ۳۶، شماره ۳، ص ۸۰-۶۷.

حسینعلی‌زاده، م.، ن. حسنعلی‌زاده، م. بابانژاد و م. رضانژاد. ۱۳۹۳. پیش بینی بارش ماهانه با استفاده از بسته‌های تخصصی سری‌های زمانی در محیط نرم افزار(R مطالعه موردی: ایستگاه آرازکوسه استان گلستان). نشریه حفاظت و بهره برداری از منابع طبیعی، جلد دوم، شماره دوم، ص ۱۲-۱.

عبداله نژاد، ک. ۱۳۹۴. مدل‌های تصادفی سری زمانی در پیش‌بینی بارندگی ماهانه. مطالعه موردی: ایستگاه هاشم آباد گرگان. مجله آمایش جغرافیایی فضا، سال پنجم، شماره مسلسل هفدهم، ص ۲۵-۱۵.

کارآموز، م. و س. عراقی‌نژاد. ۱۳۸۵. هیدرولوژی پیشرفته. ویرایش دوم. انتشارات دانشگاه امیر کبیر، تهران.

مالمیر، م. ۱۳۸۵. پیش‌بینی سری‌های زمانی کم‌آبی‌های رودخانه. دانشگاه تهران. پایان‌نامه کارشناسی ارشد در مدیریت منابع آب، دانشکده کشاورزی، ۸۰ صفحه.

Chen, K. Y. and C. H. Wang. 2007. A hybrid SARIMA and support vector machines in forecasting the production values of the machinery industry in Taiwan. Expert Systems with Applications, 32: 254-264.

Willmott, C. J., S. M. Robeson and K. Matsuura. 2011. A refined index of model performance. International Journal of Climatology, 32(13): 2088-2094.

Wang, H. R., C. Wang, X. Lin and J. Kang. 2014. An improved ARIMA model for precipitation simulations. Nonlinear Processes in Geophysics, 21:1159-1168.

Aman, M. K. 2017. Enhanced monthly precipitation forecasting using artificial neural network and singular spectrum analysis conjunction. INAE Letters, 2:73-81.

Carmen Bas, M., J. Ortiz, L. Ballesteros and S. Martorell. 2017. Evaluation of a multiple linear regression model and SARIMA model in forecasting 7Be air concentrations. Chemosphere, 177:326-333.

Dabral, P. P. and Z. Murry Mharhoni. 2017. Modeling and forecasting of rainfall time series using SARIMA. Environmental Processes, 4:399-419.

Narasimha Murthy, K. V., R. Saravana and K. V. Kumar. 2018. Modeling and forecasting rainfall patterns of southwest monsoons in North–East India as a SARIMA process. Meteorology and Atmospheric Physics, 130:99-106.

Improvement of Seasonal ARIMA Model and Investigation of Model Performance for Monthly Precipitation Simulation

Laleh Parviz[1]

Abstract

Precipitation is an important component of the hydrological cycle which links atmospheric and surficial process. Therefore, accurate modeling and estimation of parameter are needed for water resources management, irrigation scheduling, agricultural management and water allocation. SARIMA model is the most popular model for monthly precipitation simulation. The weakness of model is to ignore the inter-monthly variation within each year. Therefore, the aim of paper is the improvement of SARIMA model which takes into account the interannual, inter-monthly variation and comparison the performance of improved model with SARIMA model in Ardabil station. Ward's method for clustering of monthly precipitation time series and linear regression model for determination the relation between statistical characteristic of each cluster and monthly precipitation have been applied. 24.05%, 17.24% and 28.48% decreasing of RRMSE, RMSE and MAE from SARIMA to improved SARIMA model indicated the acceptable performance of improved model. The comparison of observed and simulated values showed the overestimation of improved model simulation. 51.16% RRMSE decreasing in seasonal precipitation simulation was expressive of accuracy increasing of improved model basis on the clustering analysis. The correlation coefficient between simulated of improved SARIMA model and observed precipitation data was increased and reached to the significant level. Therefore, the clustering analysis and improvement of SARIMA model, increased the accuracy of model.

Keywords: Precipitation, Improvement, Clustering Analysis, SARIMA.

[1] Faculty of Agriculture, Azarbaijan Shahid Madani University, Tabriz, Iran. (* Corresponding Author: laleh_parviz@yahoo.com)