نشریه پژوهش‌های اقلیم شناسی
سال یازدهم | شماره چهل و دوم | تابستان ۱۳۹۹
وصول: ۹۸/۲/۲۴ پذیرش: ۹۹/۰۱/۲۴
صص ۶۰-۴۷

افزایش دقت برآورد داده‌های گمشده بارش ماهانه با الگوریتم ژنتیک و کلونی مورچگان

محبوبه فرزندی*[۱] ، حسین ثنایی نژاد[۲]، بیژن قهرمان[۳]، مجید سرمد[۴]
۱-دانشجوی دکتری هواشناسی کشاورزی،
۲- دانشیار، گروه مهندسی آب دانشگاه فردوسی مشهد،
۳- استاد، گروه مهندسی آب دانشگاه فردوسی مشهد،
۴- دانشیار، گروه آمار، دانشگاه فردوسی مشهد

چکیده:

بارش از مهم‌ترین متغیرهای هوا و اقلیم‌شناسی بوده و ارتباط مستقیم با وضعیت اقلیمی منطقه دارد. دقت شبیه سازی این متغیر با توجه به تغییرات زیاد آن از اهمیت بسزایی برخوردار است. آمار مشاهده‌ای در اولین ایستگاه های همدید ایران از سال ۱۳۳۰ (۱۹۵۱ میلادی) در سایت سازمان هواشناسی ایران قابل دسترس است. آمار قدیمی و طولانی مدت دما و بارش ماهانه پنج شهر ایران شامل مشهد توسط سفارت امریکا و انگلیس از دوره قاجار (حدود ۱۸۸۰) اندازه‌گیری و در کتبی ثبت شده‌است. متاسفانه، این آمار دارای داده گمشده می باشد. داده های گمشده ماهانه عمدتا در طول جنگ جهانی دوم (۱۹۴۹-۱۹۴۱) و به‌طور پراکنده در طول دوره آماری وجود دارد. ایستگاه‌هایی از کشورهای مجاور با توجه به معیار فاصله، همبستگی و تکمیل بودن داده‌ها در دوره‌های دارای داده گمشده به‌عنوان ایستگاه‌های مبنا انتخاب شدند. این پژوهش ده الگوی چندگانه رگرسیونی را به بارش ماهانه ایستگاه مشهد برازش داده و سپس پارامترهای این الگوها با روش‌های الگوریتم ژنتیک و الگوریتم کلونی مورچگان بهینه کرده‌است. نتایج نشان داد الگوریتم ژنتیک و کلونی مورچگان دقت برآورد داده‌های گمشده بارش را به طور چشمگیری بالا می‌برد. کمترین معیار خطای RMSE الگوهای رگرسیونی ۹/۷۹ است که با بهینه سازی با ژنتیک الگوریتم تا ۲/۵۶۰ و با الگوریتم کلونی مورچگان تا ۲/۵۵۹ کاهش می‌یابد.

واژگان کلیدی: داده‌گواری، بارش مشهد، داده‌گمشده، رگرسیون چندگانه، روش‌های تکاملی.

* نویسنده مسئول: ۰۹۱۵۶۹۸۲۵۲۱ Email: mhb_farzandi@yahoo.com

مقدمه

الگوریتم‌های تکاملی[1] (فرگشتی) شاخه‌ای از هوش مصنوعی است که شامل مسایل بهینه‌سازی ترکیبی می‌شود. که بر اساس استفاده از قوانین داروین هستند. از لحاظ فنی این الگوریتم‌ها متعلق به حل کننده‌های آزمون و خطا هستند و می‌توان آن‌ها را از روش‌های بهینه‌سازی کلی با ماهیت فرا ابتکاری یا بهینه‌سازی تصادفی قلمداد کرد که به وسیله‌ی استفاده از جمعی از راه‌حل‌های پیشنهادی (به جای تکرار کردن یک روش در فضای جستجو) برجسته شده‌است. الگوریتم ژنتیک[2] و الگوریتم بهینه سازی کلونی مورچگان[3] از الگوریتم‌های شناخته شده تکاملی مبتنی بر جمعیت است (لیتل و روبین، ۲۰۰۲).

داده گمشده[4] بخشی از مجموعه داده هاست که گزارش نشده اند و باعث کاهش تطابق نمونه با جامعه شده و می تواند منجر به نتیجه‌گیری اشتباه در مورد جامعه اصلی شود. وجود داده گمشده در متغیرهای آب و هواشناسی یک اتفاق معمول بوده و بسته به میزان آن، می تواند اثر قابل توجهی در تحلیل‌های به دست آمده از داده ها داشته باشد. تمامی روش‌های برآورد پارامترها بر پایه فرض کامل بودن مجموعه داده‌ها استوار است و تحت برقراری این شرایط منجر به برآوردهایی نا اریب می‌شوند. با افزایش نسبت گمشدگی، مقدار اریبی افزایش می‌یابد (لیتل و روبین، ۲۰۰۲ و ارقامی و همکاران، ۱۳۸۰). حجم نمونه (طول دوره آماری) به ویژه در مناطق خشک و نیمه خشک برای تحلیل فراوانی، تحلیل سری‌های زمانی، تحلیل خشکسالی‌ها و ... باید حداقل۱۰۰ سال باشد، آن گاه می‌توان به تحلیل داده‌ها اطمینان نسبی (برآورد سازگار) پیدا کرد. دوره بازگشت نیز به طول دوره آماری وابسته است. برآورد بزرگترین دوره بازگشت با دقت قابل قبول معادل یک پنجم طول داده‌ها است (جاکوب و همکاران، ۱۹۹۹). علاوه براین پر واضح است که هرچقدر آمار طولانی‌تری از یک منطقه در اختیار باشد، درک بهتری از اقلیم منطقه و تغییرات آن در طی زمان حاصل می شود. آمار قدیمی و طولانی مدت دما و بارش ماهانه پنج شهر ایران شامل مشهد توسط سفارت امریکا و انگلیس در دوره قاجار و قبل از سال۱۳۳۰ اندازه‌گیری و در کتبی به نام World weather records ثبت شده است (موسسه اسمیتسونیان، ۱۹۲۷). بنابراین طول دوره آماری دما و بارش ماهانه این ایستگاه ها نزدیک به ۱۳۰ سال می شود و تنها ایستگاه های طولانی مدت ایران هستند. متاسفانه این آمار که می‌تواند مبنای ارزشمندی در تحقیقات و نشان دهنده تغییرات دوره ای و روند و ... باشد دارای گمشدگی بوده و تا به حال در دسترس پژوهشگران قرار نگرفته است. روش هایی کلاسیک، ابتکاری و فراابتکاری و با دقت نسبتا خوب برای برآورد داده گمشده موجود است. این پنج شهر عبارت اند از: مشهد (دما ۱۸۸۵و بارش ۱۸۹۳)، تهران (دما ۱۹۵۱و بارش ۱۸۸۴)، اصفهان (دما ۱۹۵۱ و بارش ۱۸۹۳)، جاسک (دما ۱۸۹۳ و بارش ۱۸۹۳) و بوشهر (دما ۱۸۷۸ و بارش ۱۸۷۸). تاکنون ترمیم این داده‌ها در مقیاس ماهانه (به جز دمای مشهد توسط فرزندی و همکاران، ۱۳۹۳) انجام نشده است. دو پژوهش در مقیاس سالانه بر روی داده‌های بارش این ایستگاه‌ها انجام شده است. خلیلی و بذر افشان تداوم خشکسالی‌ها را با تحلیل فراوانی بارش پنج شهر ایران با آمار طولانی مدت بررسی کردند. این پژوهشگران روش خودهمبستگی را برای ترمیم داده‌های گمشده انتخاب کردند (خلیلی و بذر افشان ،۱۳۸۷). قهرمان و احمدی پانزده سال داده های گمشده‌ی بارش سالانه بلند مدت ایستگاه مشهد را ترمیم نمودند. آنها از روش کریجینگ و برازش رگرسیون چندگانه بر میانگین‌های متحرک باران سالانه استفاده کردند و فقط اطلاعات درون خود داده‌ها را به کار بردند. ترمیم داده‌ها در پژوهش آنها در مقیاس سالانه انجام شده است (قهرمان و احمدی، ۲۰۰۷). ترمیم این آمار در قلمروی برآورد داده‌های گمشده است. الگوریتم‌های تکاملی کلونی مورچگان و الگوریتم ژنتیک به منظور افزایش دقت برآورد داده‌های گمشده بارش ماهانه مشهد و ارائه آمار واقعی طولانی مدت ماهانه برای اولین بار در ایران به کار خواهد رفت. این آمار می‌تواند مبنای ارزشمندی برای مطالعات

1 Evolutionary algorithms
2 Genetic Algorithm
3 Ant Colony Optimization
4 Missing Data

منابع آب، خشکسالی‌ها، تغییر اقلیم، گرمایش جهانی و ... باشد.

اکبال و همکاران (۲۰۱۶) تغییرات بارش در دریای زرد چین را در دوره آماری ۲۰۱۴-۱۹۶۰ بررسی کرده‌اند. آن ها روش رگرسیون خطی را برای برآورد داده های گمشده بارش به کار برده‌اند. پاتیل و بیچکار یک روش ترکیبی شامل الگوریتم ژنتیک و درخت تصمیم برای بالا بردن دقت برآورد داده‌های گمشده ارائه داده‌اند (پاتیل و بیچکار، ۲۰۱۰). اوستوریکار و دئو دو روش شبکه عصبی و برنامه-ریزی ژنتیک را برای ترمیم داده‌های گمشده ساعتی امواج در خلیج مکزیک به کار بردند. نتایج نشان داد که برنامه-ریزی ژنتیک نسبت به شبکه عصبی هنگامی که تعداد داده-های گمشده افزایش می‌یابد، برتری دارد (اوستوریکو و دئو، ۲۰۰۸). دستورانی و همکاران داده‌های گمشده‌ی ماهانه ده ایستگاه مختلف آبسنجی ایران را به چهار روش ANN، عصبی فازی، همبستگی و نسبت نرمال برآورد و نتیجه گرفتند که هر چهار روش جواب قابل قبولی را می‌دهند. روش عصبی فازی نسبت به بقیه برتر و روش شبکه عصبی در رتبه دوم قرار دارد (دستورانی و همکاران ،۲۰۱۰). یوزگاتلیجیل و همکاران شش روش ترمیم داده‌های گمشده را برای بارش و دمای ماهانه ترکیه ارزیابی و مقایسه نمودند. روش‌ها در این تحقیق به دو دسته ساده و پیچیده تقسیم شدند. روش‌های ساده عبارتند از میانگین حسابی، نسبت نرمال[1] (NR) و نسبت نرمال وزنی با روش همبستگی. روش‌های پیچیده شامل پرسپترون چند لایه شبکه عصبی مصنوعی، استراتژی جاگذاری چندگانه با زنجیره مارکوف–مونت کارلو بر اساس الگوریتم EM (EM-MCMC[2]) و یک روش اصلاح شده EM-MCMC است. آنها مجموع مربعات خطا را برای انتخاب روش برتر به کار بردند. افزون براین روش تحلیل سری‌های زمانی پویای غیرخطی نیز برای وابستگی فضایی مکانی داده‌های جاگذاری شده به کار گرفتند. تحلیل آنها نشان داد که دو روش ANN و EM-MCMC از بقیه بهتر عمل می‌کنند (یوزگاتجیل و همکاران، ۲۰۱۳). تنکالیک و همکاران داده-های گمشده‌ی دبی روزانه یک ایستگاه آبسنجی (جنوب شرقی فرانسه) را با رگرسیون و از روی ایستگاه‌های اطراف برآورد، سپس باقیمانده‌ها را با الگوی سری زمانی ARIMA تحلیل کردند. نتایج نشان داد که این ترکیب (رگرسیون پویا) دقت برآورد داده‌های گمشده را افزایش می‌دهد (تنکالیک و همکاران، ۲۰۱۵). پریس و استفیلد (۲۰۰۸) از روش الگوریتم ژنتیک برای کالیبراسیون پارامترهای مدل درخت (MT) استفاده کردند. سپس این مدل ترکیبی را برای پیش بینی اجزای کیفیت جریان آب استفاده نمودند. برخی محققین برای بالا بردن دقت برآوردهای خود از روش‌های ترکیبی استفاده نموده‌اند. لیائو و همکاران (۲۰۱۶) یک الگوریتم ابتکاری ترکیبی به نام SAEM-HMM پیشنهاد دادند که شامل مدل مخفی مارکف (HMM) الگوریتم EM و الگوریتم گرم و سرد کردن (SA) است. این الگوریتم برای غلبه بر حساسیت مدل مارکف مخفی به مقادیر اولیه از EM و برای رهایی از گیر افتادن در بهینه‌های محلی از SA استفاده می‌کند. اسعد و همکاران (۲۰۱۶) نیز روشی ترکیبی ارائه کردند که با استفاده از الگوریتم EM پارامترهای الگوی پیشنهادی را با دقت بیشتری برآورد می‌کند. الگوریتم پیشنهادی این مقاله VEM-DyMix نام دارد که از ترکیب الگوریتم تغییرات حداکثر درستنمایی (VEM[3]) با یک مدل مخلوط گوسی با متغیرهای پنهان با توزیع قدم زدن تصادفی (DyMix) به دست می آید.

مواد و روش

داده‌های گمشده یکی از مشکلات جدی در علوم مختلف به ویژه در آب و هواشناسی است. روش های ساده ای از قبیل میانگین گیری، نزدیک ترین همسایه موجود است که به دلیل دقت کم از آنها استفاده نشده است. امروزه روش-های بسیار کارآمدتری برای رفع مشکل داده های گمشده ارائه شده است که به سازوکار داده‌های گمشده بستگی دارد. ابتدا روش کلاسیک رگرسیون چندگانه به عنوان یک روش پر کاربرد در ترمیم داده های گمشده آب و

1 Normal Ratio

2 Monte Carlo Markov Chain based on expectation – maximization

3 Variational approximation EM

هواشناسی برای به دست آوردن الگوهای اولیه به کار می رود. به دلایلی از قبیل برقرار نبودن کامل فرض های پایه و پذیرش خطای اولیه در این روش از الگوریتم های تکاملی GA و ACO برای افزایش دقت پارامترهای الگوهای رگرسیونی استفاده می کنیم، که یک روش ترکیبی جدید است.

رگرسیون

تابعی است که وابستگی یک متغیر (پاسخ) به یک یا چند متغیر (پیشگو) را ارائه می‌دهد. این تابع می‌تواند خطی، غیرخطی، ساده و چندگانه باشد. رگرسیون چندگانه ابزاری سودمند در ترمیم و گسترش داده‌های گمشده است (Dingman, 2002). روش‌های دیگری نیز برای ترمیم داده‌های گمشده وجود دارد که به اطلاعات در دسترس بستگی دارد. اگر اطلاعات خوبی از ایستگاه‌های مجاور در منطقه (که همبستگی مناسبی با ایستگاه مورد نظر داشته باشند) در اختیار باشد، آنگاه رگرسیون شیوه ای مناسب برای ترمیم و گسترش داده‌هاست. همچنین رگرسیون مکملی برای برخی از روش‌های دیگر نیز می‌باشد. بررسی باقی‌مانده‌ها (آسیب‌شناسی[1]) الگوی رگرسیونی یکی از نقاط قوت رگرسیون است (رضایی‌پژند و بزرگ‌نیا، ۱۳۸۱و Ranhao, et al., 2008). شبکه عصبی در واقع نوعی رگرسیون غیرخطی است. رگرسیون امیدریاضی شرطی Y به‌شرط متغیرهای X_1 تا X_k مطابق رابطه (۱) است.

$$\mathbf{Y = E(Y \mid X_1 = x_1, \cdots, X_k = x_k)} \quad (۱)$$

رابطه (۲) رگرسیون چندگانه خطی را براساس k متغیر پیشگو نشان می‌دهد. β_0 تا β_k پارامترهای الگو هستند که باید با داده‌های در دسترس برآورد شوند. u مولفه خطاست که از توزیع نرمال با میانگین صفر و واریانس ثابت σ^2 پیروی می‌کند. σ^2 نیز باید توسط داده‌ها برآورد شود (رضائی‌پژند و بزرگ‌نیا، ۱۳۸۱ و رانهائو و همکاران، ۲۰۰۸).

$$\mathbf{Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k + u} \quad (۲)$$

قبول نهایی الگو مشروط به برقراری فرض های پایه است.

۱) خطاها توزیع نرمال دارند. (آزمون شاپیرو)

۲) میانگین خطاها صفر است.

۳) واریانس خطاها یکسان و ثابت است. (پایایی واریانس)

۴) خطاها به طور ناخودهمبسته توزیع شده‌اند. (دربین واتسون)

۵) داده پرت کنترل و بررسی می‌شود. (آماره کوک)

الگوریتم ژنتیک (GA)

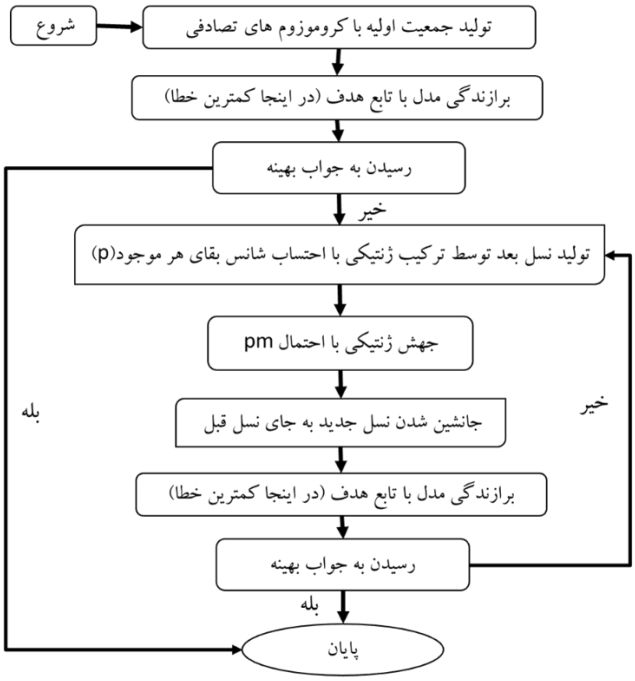
الگوریتم ژنتیک روش جستجوی احتمالاتی است که از فرآیند تکامل زیست شناختی طبیعی پیروی می‌کند. این روش بر اساس نظریه تکاملی داروین بنا شده و جواب مساله‌ای که از این طریق به دست می‌آید، در طول فرآیند تولید جواب، بهبود می‌یابد. GA با یک مجموعه از جواب‌ها که با کروموزوم‌ها نشان داده می‌شوند شروع می‌شود. این مجموعه جواب ها جمعیت[2] اولیه نام دارد. جواب‌های حاصل از یک جمعیت برای تولید جمعیت بعدی استفاده می‌شوند. معمولا جمعیت جدید نسبت به جمعیت قبلی در این فرآیند بهتر است. انتخاب بعضی از جواب‌ها از میان کل جواب‌ها (والدین)[3] به منظور ایجاد جواب‌های جدید یا همان فرزندان[4] بر اساس میزان برازندگی آن‌ها است. طبیعی است که جواب‌های مناسب‌تر شانس بیشتری برای تولید دوباره دارند. این فرآیند تا برقراری شرط تعیین شده (مانند تعداد جمعیت‌ها یا میزان بهبود جواب) ادامه می‌یابد. دو عملگر ترکیب و جهش پایه‌ی الگوریتم ژنتیک است. روش‌های مختلف ترکیب والدین به منظور ایجاد فرزندان (والدین جدید) وجود دارد. اما انتخاب والدین بهتر ایده اصلی در تمام آن‌ها است، به این امید که تولید فرزندان بهتر شود. اگر جمعیت جدید تنها از طریق فرزندان جدید ایجاد شود، این فرآیند منجر به حذف کروموزوم‌های نسل قبل می‌شود. همیشه بهترین جواب نسل قبل برای جلوگیری از این پیشامد (بدون هیچ تغییری) به نسل جدید منتقل می‌شود (آیدلینک و آرسنال، ۲۰۱۳).

1 Diagnostic

2 population
3 parents
4 offspring

اعضای موجود در جامعه یا به عبارت دیگر تقریب‌های جاری جواب بهینه مسئله، هر کدام به صورت رشته‌ای همانند کروموزوم با پشت سرهم قرار گرفتن متغیرهای مساله کدگذاری می‌شوند. تابع هدف، شاخصی برای انتخاب جفت های مناسب و جفت‌گیری آنها را در جمعیت کروموزوم ها فراهم می‌کند. پس از تولید نسل اولیه، نسل-های بعدی با عملیات ژنتیک مانند نخبه‌گرائی، جفت‌گیری (تقاطع) و جهش تولید می‌شود. به هر عضو جمعیت ارزشی نمایانگر میزان سازگاری آن با تابع هدف تعلق می‌گیرد. اعضایی که سازگاری بیشتری دارند شانس بالاتری برای انتخاب و انتقال به نسل بعد خواهند داشت (نخبه‌گرایی). عملگر جفت‌گیری ژنتیک، ژن‌های کروموزوم‌ها را با یکدیگر ترکیب می‌کند. اما لزوما بر تمام رشته‌های جمعیت عمل نمی‌کند. بالعکس اعمال آن با احتمال p صورت می-پذیرد. عملگر جهش با انتخاب تصادفی چند کروموزوم و انتخاب تصادفی یک یا چند ژن و جانشینی نقیض آن انجام می‌شود. این بار نیز عمل جهش با احتمال معین pm صورت می پذیرد. به این صورت چرخه فرآیند در نسل های بعدی ادامه می‌یابد. پایان فرآیند الگوریتم رسیدن به جواب های بهینه مطلوب است (سیدنژاد، ۱۳۹۱). با توجه به توضیحات فوق فلوچارت عملیاتی الگوریتم ژنتیک تهیه و تنظیم شد که در شکل (۱) آمده است.

الگوریتم ژنتیک برای برای برآورد و بهینه‌سازی پارامترهای الگوهای رگرسیونی به کار می‌رود. تابع هدف کمینه سازی جذر میانگین خطا (RMSE) است. تابع خطا با تکرار الگوریتم و تولید پارامترهای بهتر در هر نسل بهینه می‌شود.



شکل ۱- فلوچارت کلی مراحل انجام الگوریتم ژنتیک.

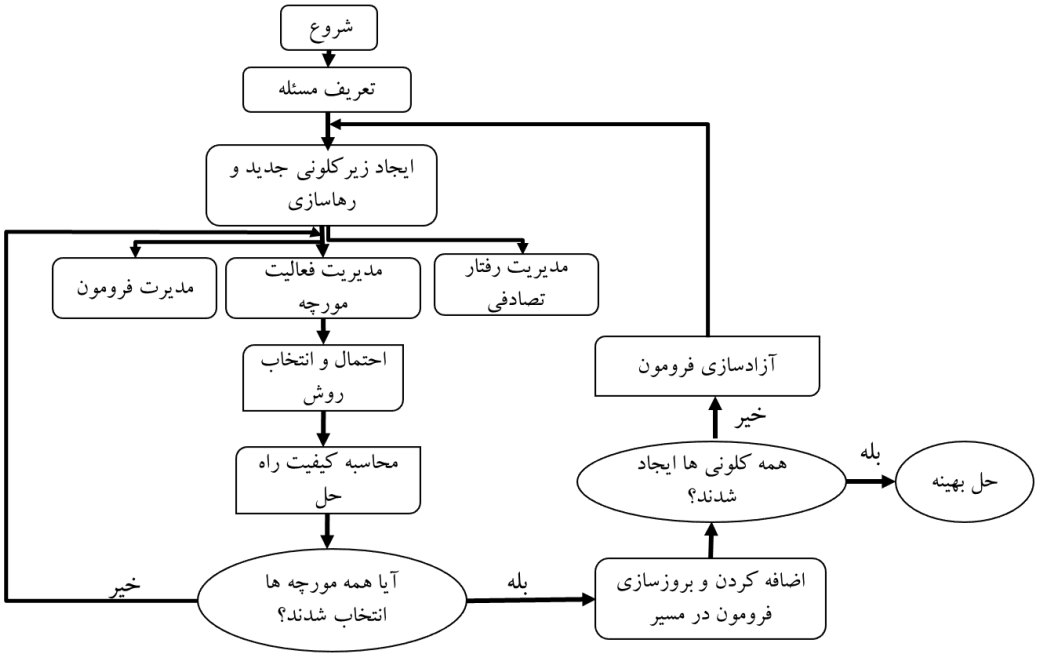
بهینه‌سازی گروه مورچه‌ها (ACO)

الگوریتم کلونی مورچگان یکی از بارزترین نمونه‌ها برای روش‌های هوش جمعی است. این الگوریتم از روی رفتار جمعی مورچه‌ها الهام گرفته شده است. مورچه ها با همکاری یکدیگر، کوتاه‌ترین مسیر را میان لانه و منابع غذایی پیدا می کنند تا بتوانند در کمترین زمان مواد غذایی را به لانه منتقل کنند. هیچ کدام از مورچه ها، به تنهایی قادر به انجام چنین کاری نیستند، اما با همکاری و پیروی از چند اصل ساده، بهترین راه را پیدا می‌کنند. مورچه‌ها مسیرهای رسیدن به غذا را به تصادف انتخاب می‌کنند. هر مورچه ماده

شیمیایی (به نام فرومون) را در هنگام حرکت به عنوان اثر روی مسیر به جا می‌گذارد. تعداد ترددهای زیاد و ایجاد فرومون بیشتر منجر به ایجاد مسیر بهینه می‌شود. تبخیر شدن فرومون و احتمال-تصادف به مورچه‌ها امکان پیدا کردن کوتاه‌ترین مسیر را می‌دهد. این دو ویژگی باعث ایجاد انعطاف در حل هرگونه مسئله بهینه‌سازی می‌شوند. ACO بهترین مسیر را در یک تابع با این خاصیت و راه حل بهتری برای مساله با محاسباتی-عددی بر مبنای علم احتمالات پیدا می‌کند. هر مورچه که ماده شیمیایی را در مسیر خود به جا می‌گذارد با احتمال رابطه (۳) می تواند مورچه بعدی را در این مسیر قرار دهد.

$$P_A(t+1)=\frac{(c+n_A(t))^{\alpha}}{(c+n_A(t))^{\alpha}+(c+n_B)^{\alpha}}=1-P_B(t+1) \quad (۳)$$

$n_A(t)$ و $n_B(t)$ تعداد مورچه هایی که در زمان t در مسیر A و B قرار دارند. c درجه جذب برای یک مسیر ناشناخته است. هر چه c بزرگتر باشد به معنی مقدار فرومون بیشتر برای انتخاب مسیر است. α و c پارامترهای مدل می‌باشند (چادهوری و همکاران، ۲۰۱۴). با توجه به توضیحات فوق فلوچارت عملیاتی الگوریتم مورچه ها تهیه و تنظیم شد که در شکل (۲) آمده است.



شکل ۲- فلوچارت عملیاتی که در اجرای الگوریتم مورچه ها انجام می شود.

نتایج و بحث

الگوهای رگرسیونی به داده های بارش بدون مشاهدات گمشده برازش داده می شود و سپس با ارزیابی مدل ها، چنان چه مدل مناسبی به دست آمد از آن می توانیم برای پیش بینی داده های گمشده در ایران استفاده نماییم. سپس روش های تکاملی به دلیل مناسب بودن برای داده هایی با نوسان زیاد شبیه بارش و نداشتن محدودیت های روش های کلاسیک به منظور بالابردن دقت الگوها به کار می رود.

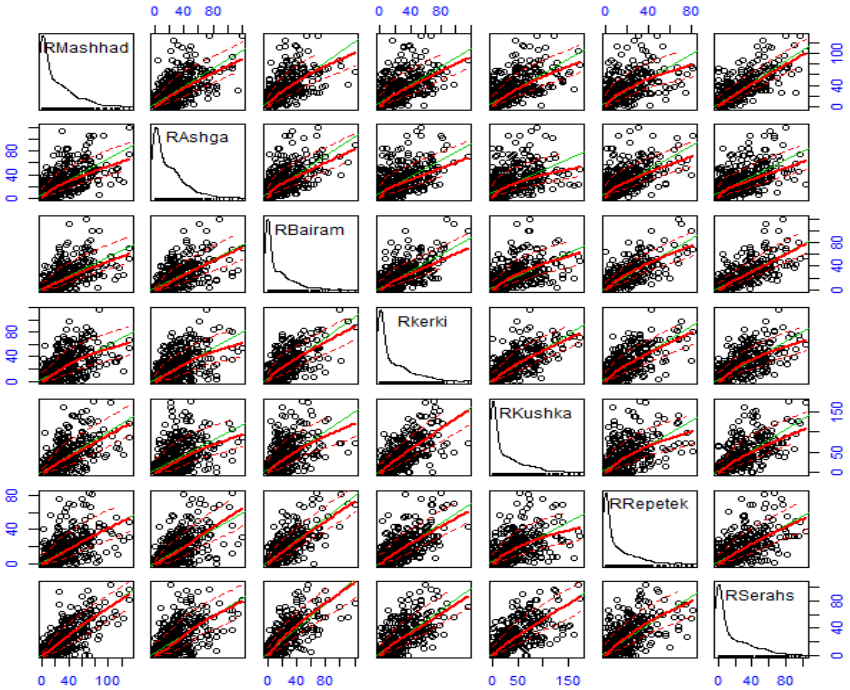
بارش ایستگاه های عشق آباد (R_{Ash}) از کشور تاجیکستان و سرخس (R_{Ser})، کوشکا (R_{Kus})، بایرام علی (R_{Bai})، کرکی (R_{Ker}) و ریتک (R_{Rep}) از کشور ترکمنستان به عنوان متغیر مستقل در ساخت داده‌های گمشده بارش مشهد (R_{Mas}) انتخاب شدند. سه عامل فاصله تا ایستگاه مشهد، همبستگی و وجود داده در ماه های گمشده در انتخاب این ایستگاه ها موثر بودند. نام ایستگاه‌های برگزیده، فاصله تا ایستگاه مشهد و همبستگی با بارش این ایستگاه در جدول (۱) مشخص شده است. آزمون t در کلیه موارد (با p-value < 2.2e-16)

معنی داری همبستگی پیرسون بارش ماهانه شش ایستگاه برگزیده با ایستگاه پاسخ را تایید می‌کند.

جدول۱- اطلاعات مربوط به ایستگاه های منتخب بارشی از نظر همبستگی و فاصله با ایستگاه مشهد.

ایستگاه‌ها	مشهد	سرخس	عشق آباد	کوشکا	بایرام علی	رپتک	کرکی
فاصله تا مشهد (mil)		۹۱/۴۸	۱۳۷/۴۳	۱۶۶/۶۸	۱۶۸/۸۵	۲۵۲/۲۸	۳۲۶/۴۶
فاصله تا مشهد (km)		۱۴۷/۲	۲۲۱/۱	۲۶۸/۲	۲۷۱/۷	۴۰۵/۹	۵۲۵/۳
آزمون همبستگی	همبستگی	۰/۸۱	۰/۷۰	۰/۷۰	۰/۶۸	۰/۶۸	۰/۶۴
	آماره t	۴۱/۴۱	۳۲/۰۸	۳۰/۶۸	۳۰/۹۵	۲۴/۸۰	۲۷/۵۵

لازم به ذکر است داده‌ها به صورت ماهانه وارد شده است تا اثر افزایشی سری زمانی را بر معنی‌داری همبستگی داده ها را کم کند. تعداد زیاد داده‌ها در معنی‌داری ضریب همبستگی تاثیر داشته است، بنابراین نمودار پراکنش بارش ایستگاه‌ها در برابر یکدیگر نیز ترسیم شد. نمودار ماتریسی بارش مشهد در برابر بارش ایستگاه‌های مستقل در شکل (۳) آمده است. شکل (۳) نشان دهنده همبستگی مثبت داده-های بارشی در برابر یکدیگر است.



شکل ۳- رسم نمودار پراکنش ماتریسی بارش ماهانه ایستگاه مشهد در برابر بارش ایستگاه های مستقل

الگوهای رگرسیونی

ترمیم بارش ماهانه مشهد به کمک ایستگاه های ذکر شده پنج الگوی رگرسیونی خطی چندگانه (P1، P3، P5، P7 و P9) با الگوهای رگرسیونی خطی چندگانه با تغییر متغیرهای مستقل با لگاریتم گیری (P2، P4، P6، P8 و P10) به شرح زیر انجام شده است. چون لگاریتم برای اعداد حقیقی مثبت تعریف شده است و بارش ماهانه حداقل برابر صفر است بنابراین برای جلوگیری از این خطا عدد یک را به متغیر بارش داخل لگاریتم اضافه نمودیم. الگوهای مختلف برای مقایسه و انتخاب برترین الگو برازش داده شد. این

الگوبندی ها با برنامه نویسی در محیط R-Studio انجام شد. متغیرهای بارش ماهانه این هفت ایستگاه را با علائم اختصاری (بخش قبل) تعریف می‌کنیم. الگوهای برازشی به ترتیب در روابط با شماره (P۱) تا (P۱۰) آمده است. نتایج حاصل از این الگوها شامل ضرایب، عامل تورم واریانس (VIF)، آماره دوربین–واتسون، خطا و قدرت الگو در جدول (۲) آمده است.

الگوی (P۱) $R_{Mas} = \beta_0 + \beta_1 R_{Ash} + \beta_2 R_{Kus} + \beta_3 R_{Ser}$

الگوی(P۲) $log(R_{Mas} + 1) = \beta_0 + \beta_1 log(R_{Ash} + 1) + \beta_2 log(R_{Kus} + 1) + \beta_3 log(R_{Ser} + 1)$

الگوی(P۳) $R_{Mas} = \beta_0 + \beta_1 R_{Kus} + \beta_2 R_{Rep} + \beta_3 R_{Ser}$

الگوی (P۴) $log(R_{Mas} + 1) = \beta_0 + \beta_1 log(R_{Kus} + 1) + \beta_2 log(R_{Rep} + 1) + \beta_3 (R_{Ser+1})$

الگوی(P۵) $R_M = \beta_0 + \beta_1 R_{Bas} + \beta_2 R_{Kus} + \beta_3 R_{Ser}$

الگوی (P۶) $log(R_{Mas} + 1) = \beta_0 + \beta_1 log(R_{Bas} + 1) + \beta_2 log(R_{Kus} + 1) + \beta_3 log(R_{Ser} + 1)$

الگوی (P۷) $R_{Mas} = \beta_0 + \beta_1 R_{Kus} + \beta_2 R_{Ker} + \beta_3 R_{Kus}$

الگوی (P۸) $log(R_{Mas} + 1) = \beta_0 + \beta_1 log(R_{Bas} + 1) + \beta_2 log(R_{Ker} + 1) + \beta_3 log(R_{Kus} + 1)$

الگوی (P۹) $R_{Mas} = \beta_0 + \beta_1 R_{Ash} + \beta_2 R_{Bas} + \beta_3 R_{Ser}$

الگوی (P۱۰) $log(R_{Mas} + 1) = \beta_0 + \beta_1 log(R_{Ash} + 1) + \beta_2 log(R_{Bas} + 1) + \beta_3 log(R_{Ser} + 1)$

آزمون‌های مناسب برای تحلیل باقیمانده ها (مباحث تشخیصی الگو) انجام شد. این آزمون ها برقراری فرض های زیربنایی رگرسیون را بررسی می کند. خلاصه نتایج الگوهای (P۱) تا (P۱۰) در جدول (۲) آمده است.

بررسی ناخودهمبستگی خطا ها: آماره آزمون دربین–واتسون در محدوده ناهمبسته بودن مقادیر جدول دوربین–واتسون (۲/۲۶– ۱/۶۰) قرار دارد. بنابراین باقیمانده ها ناخودهمبسته مرتبه اول اند (جدول ۲). شرح مختصر بقیه موارد مباحث تشخیصی الگوها در ادامه آمده است. جداول و نمودارهای آن به دلیل حجم بالا حذف شده است.

بررسی همگن بودن واریانس خطا: مقدار آماره کای–دو در آزمون بروش–پاگن (دستور NCV test از بسته نرم افزاری car در نرم افزار R) با مقادیر احتمال نزدیک به صفر نشان دهنده مشکوک بودن فرض همگن بودن واریانس باقی-مانده‌هاست. هرچند نمودار پراکنش تثبیت واریانس را تایید می‌کند (مقادیر پیش بینی شده در برابر باقی‌مانده های استاندارد شده به صورت ایده آل و مستطیلی توزیع شده-اند).

بررسی داده پرت: مقدار کم آماره میانگین فاصله کوک در تمام الگوها (۰/۰۰۱۲ تا ۰/۰۰۲۴) عدم وجود داده پرت را نشان می‌دهد.

بررسی نرمال بودن خطاها: بررسی نرمال بودن باقی‌مانده‌ها با آزمون شاپیرو و نمودار چندکی انجام می‌شود. نتایج نشان داد باقی‌مانده‌ها در الگوهای بارشی کمی از حالت نرمال منحرف شده‌اند. با توجه به اینکه اندازه نمونه به میزان کافی بزرگ و سایر فرض‌های کلاسیک برقرار است، انحراف از فرض نرمال بودن باقی‌مانده ها معمولا کم اهمیت و پیامدهای آن ناچیز است (علی سوری، ۱۳۹۶) لذا از آن صرف نظر شد.

جدول ۲- ضرایب الگوهای بارش مشهد، عامل تورم واریانس (VIF)، آماره دربین-واتسون، خطا و قدرت الگو

ضرایب		برآورد ضریب	خطای معیار	آماره t	همخطی (VIF)	ضریب تعیین (R^2_{adj})	جذر مربع خطا (RMSE)	آماره F	دوربین-واتسون (D-W)
	β_0	۲/۰۵	۰/۵۳	۳/۸۷	-				
الگوی (P۱)	β_1	۰/۲۹	۰/۰۳	۱۱/۳۴	۱/۸	۰/۷۸	۹/۷۹	۹۶۸	۱/۷۶
	β_2	۰/۱۱	۰/۰۲	۵/۹۲	۲/۵				
	β_3	۰/۶۸	۰/۰۳	۱۹/۷۴	۳/۱				
	β_0	۰/۲۷	۰/۰۴	۶/۲۷	-				
الگوی (P۲)	β_1	۰/۳۴	۰/۰۳	۱۲/۹۶	۲/۴	۰/۸۱	۱۳/۵۶	۱۱۶۵	۱/۷۷
	β_2	۰/۱۷	۰/۰۳	۵/۵۷	۵/۱				
	β_3	۰/۴۶	۰/۰۴	۱۲/۵۹	۵/۷				
	β_0	۳/۸۴	۰/۵۷	۶/۷۶	-				
الگوی (P۳)	β_1	۰/۱۱	۰/۰۲	۴/۶۳	۲/۸	۰/۷۵	۱۲/۰۰	۶۷۷	۱/۶۱
	β_2	۰/۲۲	۰/۰۵	۴/۷۳	۲/۳				
	β_3	۰/۷۳	۰/۰۴	۱۸/۴۶	۳/۲				
	β_0	۰/۵۴	۰/۰۴	۱۳/۰۹	-				
الگوی (P۴)	β_1	۰/۱۹	۰/۰۴	۴/۵۰	۶/۸	۰/۷۹	۱۴/۳۱	۸۲۴	۱/۶۰
	β_2	۰/۱۹	۰/۰۴	۴/۴۰	۴/۷				
	β_3	۰/۵۲	۰/۰۵	۱۱/۱۶	۷/۲				
	β_0	۴/۱۵	۰/۵۳	۷/۸۲	-				
الگوی (P۵)	β_1	۰/۰۹۶	۰/۰۳۸	۲/۵۰	۲/۸	۰/۷۴	۱۲/۲۳	۸۰۱	۱/۶۵
	β_2	۰/۱۴	۰/۰۲	۶/۶۷	۲/۵				
	β_3	۰/۷۸	۰/۰۴۲	۱۸/۵۷	۳/۹				
	β_0	۰/۵۴	۰/۰۳۷	۱۴/۶۸	-				
الگوی (P۶)	β_1	۰/۳۰	۰/۰۴۱	۷/۳۹	۶/۳	۰/۷۹	۱۴/۸۰	۱۰۷۵	۱/۶۷
	β_2	۰/۱۹	۰/۰۳۲	۵/۸۲	۵/۴				
	β_3	۰/۴۲	۰/۰۴۵	۹/۳۴	۸/۲				
	β_0	۴/۹۱	۰/۵۹	۸/۳۳	-				
الگوی (P۷)	β_1	۰/۴۱	۰/۰۳۹	۱۰/۵۵	۲/۴	۰/۶۳	۱۳/۴۸	۴۹۷	۱/۶۰
	β_2	۰/۲۲	۰/۰۴۱	۵/۳۹	۳/۱				
	β_3	۰/۲۶	۰/۰۲۲	۱۱/۶۲	۲/۳				
	β_0	۰/۵۷	۰/۰۳۹	۱۴/۵۶	-				
الگوی (P۸)	β_1	۰/۴۷	۰/۰۳۷	۱۲/۴۶	۴/۷	۰/۷۵	۱۵/۸۴	۸۹۳	۱/۷۰
	β_2	۰/۱۱	۰/۰۳۸	۲/۸۹	۵/۵				
	β_3	۰/۲۹	۰/۰۳۱	۹/۲۳	۴/۸				
	β_0	۲/۸۳	۰/۵۵	۵/۱۵	-				
الگوی (P۹)	β_1	۰/۲۸	۰/۰۲۸	۹/۸۹	۲/۰۱	۰/۷۴	۱۱/۴۸	۸۰۶	۱/۷۱
	β_2	۰/۰۴	۰/۰۳۹	۰/۹۸	۲/۹				
	β_3	۰/۷۸	۰/۰۳۶	۰/۳۳	۲/۹				
	β_0	۰/۲۹	۰/۰۴۲	۶/۸۴	-				
الگوی (P۱۰)	β_1	۰/۲۱	۰/۰۴۱	۵/۱۹	۶/۹	۰/۸۱	۱۲/۳۲	۱۲۲۶	۱/۷۸
	β_2	۰/۳۰	۰/۰۲۷	۱۱/۲۲	۱/۸				
	β_3	۰/۴۷	۰/۰۳۷	۱۲/۷۴	۶/۲				

بهینه سازی الگوها با GA و ACO

برآورد پارامترهای رگرسیونی به روش کمترین مربعات خطا انجام می شود. چون در روش رگرسیون کمترین مربعات اشکالاتی وجود دارد: ۱- خطایی به اندازه α=۰۵/۰ را در این روش می پذیریم. پس حداقل خطایی وجود دارد. ۲- فرض های پایه دقیقا صادق نیست پس پارامترها از دقت لازم برخوردار نیست. ۳- ضریب عدم تعیین ($1-R^2$) هرچه بیشتر باشد امکان کاهش خطا بیشتر است. ۴- پارامترهای برآورد شده فاصله اطمینان دارند پس در این فاصله می توانند تغییر کنند. بنابراین می توان دقت این روش را بالا برد. این پژوهش روشی را برای ارتقای افزایش دقت الگو با دو روش بهینه سازی GA و ACO و به عنوان نوآوری ارائه و نتایج جالبی به دست آمده است. صورت کلی الگوها را می پذیریم سپس با روش های بهینه سازی GA و ACO پارامترهای الگوها را برآورد می‌کنیم. با توجه به اینکه نتایج الگوهای بارشی مشابه هستند پنج الگوی اصلی (الگوهای خطی) انتخاب شدند. پارامترهای این پنج الگو (جدول ۳) به کمک GA و ACO بهینه‌سازی می‌شوند. شبیه سازی این الگوریتم ها در محیط نرم افزار متلب ۲۰۱۷ انجام شده-است. برازش GA نیاز به پیش فرض‌هایی دارد. محدوده تغییرات ضرایب بین ۲۰- تا ۲۰ انتخاب شد. این انتخاب بر اساس تحلیل پایلوت (الگوهای رگرسیونی مبنا) به دست آمد. بیشترین تکرار الگوریتم از ۲۰۰ تا ۳۰۰۰ برای هر الگو متغیر است. ورودی های اولیه مشترک برای الگوها مطابق جدول (۳) است. درصد تقاطع معادل ۰/۸ یعنی ۸۰٪ داده ها به عنوان والدین برای نسل بعد انتخاب شدند. احتمال جهش نشان می‌دهد ۳۰٪ داده ها برای جهش انتخاب شد (سیدنژاد، ۱۳۹۱). اجرای کل الگوریتم برای هر الگو ۲۰ بار تکرار شد و بهترین جواب‌ها (کمترین خطا) انتخاب شد.

بهینه سازی با ACO نیز نیازمند مقادیر اولیه است. پارامترهای ثابت در همه الگوها در جدول (۴) آمده است. تعداد جمعیت اولیه را ۱۰ در نظر می گیریم. مقدار وزن ها و احتمالات به دست آمده مطابق جدول (۵) است. ACO ذاتا برای مسائل گسسته طراحی شده است، چون ما با اعداد حقیقی سروکار داریم راهکار ارائه شده در نظر گرفتن توابع گوسی بر مبنای جمعیت اولیه است. زیتا پارامتری شبیه انحراف معیار برای تقریب تابع گوسی است (سیدنژاد، ۱۳۹۴). تعداد جمعیت اولیه با آزمون و خطا ۱۰ در نظر گرفته می‌شود.

جدول ۳- مفروضات و ورودی های اولیه در همه الگوها برای اجرای GA

روش	پارامتر μ	پارامتر گاما	تعداد والدین	احتمال[1] μ	درصد تقاطع	تعداد جمعیت اولیه[2]	بیشترین تکرار[3]
چرخ رولت[4]	۰/۰۲	۰/۰۵	۲	۰/۳	۰/۸	۳۰	۱۰۰۰ تا ۳۰۰۰

جدول ۴- پارامترهای ثابت در همه الگوها برای بهینه‌سازی الگوریتم ACO

نرخ فاصله-خطا (زیتا)	فاکتور شدت q	اندازه نمونه n	بیشترین تکرار	تعداد جمعیت اولیه
۱	۰/۵	۵۰	۱۰۰	۱۰

جدول ۵- مقدار وزن ها و احتمالات به دست آمده در همه الگوهای ACO

وزن ها (w)	۰/۰۷۹۸	۰/۰۷۸۲	۰/۰۷۳۷	۰/۰۶۶۶	۰/۰۵۷۹	۰/۰۴۸۴	۰/۰۳۸۸	۰/۰۲۹۹	۰/۰۲۲۲	۰/۰۱۵۸
احتمالات (p)	۰/۱۵۶۰	۰/۱۵۲۹	۰/۱۴۴۰	۰/۱۳۰۳	۰/۱۱۳۳	۰/۰۹۴۶	۰/۰۷۵۹	۰/۰۵۸۶	۰/۰۴۳۴	۰/۰۳۰۹

[1] درصد انتخاب داده

[2] npop

[3] max Iteration

[4] Roulette wheel

نتایج برازش الگوریتم‌های GA و ACO شامل ضرایب الگوهای بهینه شده، بیشترین تکرار در الگوریتم برای رسیدن به مقدار بهینه و معیار خطا (RMSE) برای هر الگو مطابق جدول (۷) است. ستون های ششم و هشتم جدول مربوط به خطای برآورد برای دو روش تکاملی الگوریتم ژنتیک و کلونی مورچه ها ($RMSE_{GA}$ و $RMSE_{ACO}$) است. همان‌طور که در جدول آمده است RMSE با روش های تکاملی در همه الگوها کمتر از ۳ میلی متر است. بنابراین این روش‌ها در مقایسه با روش‌های رگرسیونی در برآورد بارش ماهانه بسیار موفق عمل کرده و خطای الگوها به طور چشمگیری کاهش یافته است. الگوی (۱) بارش پس از بهینه‌سازی به کمک GA و ACO کمترین خطای RMSE را دارد (جدول۶). لذا داده‌های گمشده ماهانه بارش طولانی مدت مشهد با یکی از این الگوها تکمیل و جاگذاری می گردد.

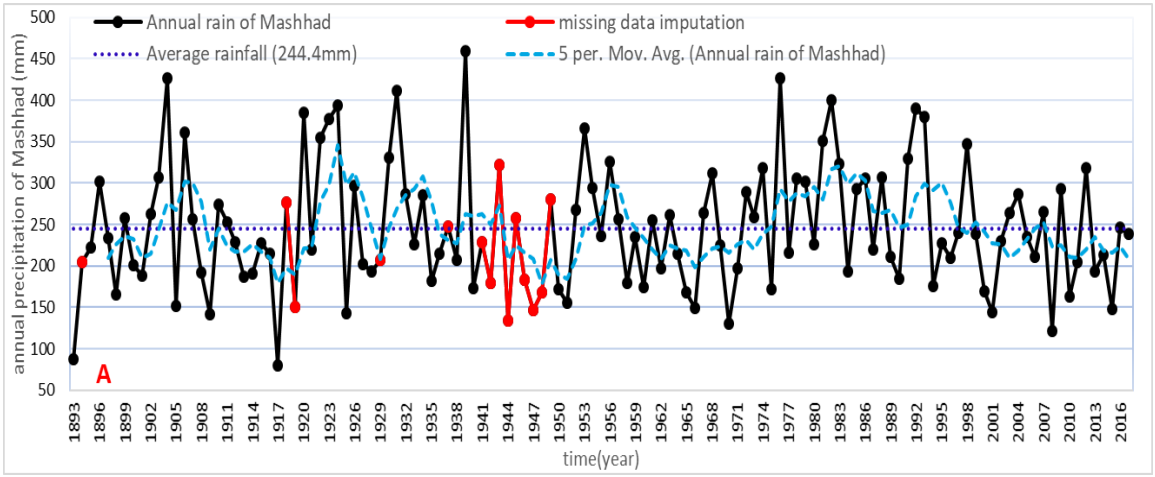
جدول۶- جدول ضرایب اصلاح شده پنج الگوی انتخابی بارش با الگوریتم ژنتیک و الگوریتم کلونی مورچگان

		برآورد ضریب GA	بیشترین تکرار[۱]	NFE	$RMSE_{GA}$	برآورد ضریب ACO	$RMSE_{ACO}$
الگوی (P۱)	β_0	-۰/۰۰۲۶	۱۰۰۰	۳۳۰۳۰	۲/۵۶۰	-۲/۷۶۸	۲/۵۵۹
	β_1	۰/۲۷۶				$\times 10^{-7}$	
	β_2	۰/۱۲۳				۰/۲۶۲	
	β_3	۰/۶۶۵				۰/۱۳۷	
الگوی (P۳)	β_0	۰/۱۲۲	۲۰۰۰	۶۶۰۳۰	۲/۶۲۸	۰/۶۱۸ × ۱۰$^{-10}$ ۴	۲/۶۲۸
	β_1	۰/۱۰۱				۰/۱۱۹	
	β_2	۰/۳۲۵				۰/۳۱۲	
	β_3	۰/۷۱۷				۰/۷۰۲	
الگوی (P۵)	β_0	۰/۰۰۰۲۸	۳۰۰۰	۹۹۰۳۰	۲/۶۷۳	۲/۳۰۴ × ۱۰$^{-6}$	۲/۶۷۳
	β_1	۰/۱۱۶				۰/۱۴۶	
	β_2	۰/۱۴۸				۰/۱۶۸	
	β_3	۰/۷۹۴				۰/۷۲۶	
الگوی (P۷)	β_0	۰/۰۰۰۶۵	۳۰۰۰	۹۹۰۳۰	۲/۹۲۵	۱/۷۵۳ × ۱۰$^{-6}$	۲/۹۲۵
	β_1	۰/۴۸۵				۰/۴۴۳	
	β_2	۰/۲۳۵				۰/۲۸۶	
	β_3	۰/۲۵۵				۰/۲۳۷	
الگوی (P۹)	β_0	۰/۰۰۰۵۵	۳۰۰۰	۹۹۰۳۰	۲/۶۴۹	۶/۲۵۱ × ۱۰$^{-9}$	۲/۶۴۸
	β_1	۰/۲۸۰				۰/۲۹۶	
	β_2	۰/۰۹۲				۰/۰۳۲	
	β_3	۰/۷۷۵				۰/۸۲۰	

۱ max Iteration

مقایسه روش کلاسیک رگرسیون و الگوریتم های تکاملی

الگوهای رگرسیونی به عنوان روش کلاسیک و بهینه‌سازی با الگوریتم های GA و ACO به عنوان روش های تکاملی در این مقاله برای ترمیم داده‌های گمشده بارش طولانی مدت مشهد بررسی و مقایسه شدند. با توجه به نتایج الگوهای رگرسیونی که در جدول (۳) آمده است کمترین معیار خطای (RMSE) مربوط به الگوی (P1) با مقدار ۹/۷۹ میلی‌متر می‌باشد. جدول (۷) که نتایج الگوها پس از بهینه سازی پارامترها با GA و ACO است، نشان می‌دهد معیار RMSE با روش GA به ۲/۵۶۰ میلی‌متر و با ACO به ۲/۵۵۹ کاهش می‌یابد. جدول (۷) همچنین نشان می‌دهد میزان خطای دو روش تکاملی تفاوت چندانی ندارد. پس از بررسی فیزیکی داده های برآوردی با این الگوریتم‌ها معلوم شد روش ACO مقادیر گمشده را کم برآورد می‌کند لذا روش GA به عنوان روش برتر انتخاب شد. نمودار سری زمانی بارش طولانی مدت سالانه مشهد پس از تکمیل و ترمیم با الگوی برتر الگوریتم ژنتیک در شکل (۳) آمده-است.



شکل۳- سری زمانی ۱۲۵ ساله بارش مشهد پس از ترمیم و تکمیل داده های ماهانه با روش بهینه سازی با الگوریتم ژنتیک، نقاط سیاه مربوط به بارش سالانه دراز مدت، نقاط قرمز مربوط به جاگذاری داده های گمشده با روش تکاملی GA، میانگین متحرک ۵ ساله با خط چین آبی و میانگین داده ها با خط چین سرمه ای نشان داده شده است.

نتیجه گیری

جانهی مقادیر از دست رفته پارامترهای هواشناسی با کمترین خطای ممکن همواره از اهمیت بالایی برخوردار بوده است. این اهمیت در مورد پارامتر بارش به دلیل نوسانات زیاد و همبستگی کمتر با سایر ایستگاه ها دوچندان شده است. لذا پرداختن به روش‌های بهینه‌سازی تکاملی به-جای روشهای معمول ضروری به نظر می‌رسد. الگوریتم ژنتیک و الگوریتم کلونی مورچگان در این تحقیق به‌منظور بالا بردن دقت شبیه‌سازی مقادیر گمشده بارش ۱۲۵ساله ماهانه مشهد به کار رفته است. نتایج نشان داد این روش های بهینه سازی مقدار خطای RMSE را به طور قابل ملاحظه ای کاهش می دهد (مقایسه خطا در جداول ۳ و۵). در بهترین حالت در الگوی (۱) از ۹/۷۹ با روش رگرسیونی به ۲/۵۶ میلیمتر با الگوریتم ژنتیک و ۲/۵۵ با الگوریتم کلونی مورچگان می رسد. این نتیجه با کاربرد روش‌های تکاملی سازگار است. زیرا اصلی ترین انگیزه استفاده از الگوریتم های تکاملی در داده کاوی ویژگی‌های جذاب آنها است که به آنها امکان می‌دهد برخی از اشکالات را در تکنیک های داده کاوی معمولی را برطرف کرده و آنها را قادر به کشف راه حل‌های جدید مانند استحکام هنگام برخورد با داده های پر پرآشوب و توانایی تفسیر داده ها بدون دانش قبلی است (عباس و همکاران، ۲۰۰۲).

شکل (۳) سری زمانی بارش دراز مدت سالانه مشهد را پس از تکمیل داده ها برای اولین بار نشان می دهد. این آمار می‌تواند مبنای ارزشمندی برای تحقیقات آب و هواشناسی باشد.

سپاس گزاری

از جناب استاد **حجت رضایی پژند** به جهت راهنمایی های ارزنده شان در انجام این پژوهش بسیار متشکریم.

منابع

۱. ارقامی، ن.ر.، سنجری، ن.، بزرگ نیا، الف، (۱۳۸۰): *مقدمه ای بر بررسی های نمونه ای*، چاپ چهارم، ۴۳۵ صفحه.

۲. خلیلی، ع.، بذرافشان، ج.، (۱۳۸۷): ارزیابی مخاطره تداوم خشک سالی با استفاده از داده های بارندگی سالانه قرن گذشته در ایستگاه های قدیمی ایران، *مجله ژئوفیزیک ایران*، جلد ۲، شماره ۲.

۳. رضائی‌پژند. حجت. بزرگ نیا. ابوالقاسم. (۱۳۸۱): *تحلیل رگرسیون غیرخطی و کاربردهای آن*. انتشارات دانشگاه فردوسی مشهد. ۳۹۸ صفحه.

۴. سوری، ع.، (۱۳۹۶): *اقتصادسنجی (پیشرفته) ج ۲ همراه با کاربرد stata12 و eviews8* ، انتشارات فرهنگ شناسی، ۱۰۲۲ صفحه.

۵. سیدنژاد گلخطمی، ن.، ۱۳۹۴. کاربرد روش های فراابتکاری در تعیین بهینه الگوهای هواشناسی. رساله دوره کارشناسی ارشد، دانشگاه پیام نور مشهد.

۶. فرزندی، م.، رضایی پژند, ح.، ثنائی نژاد، ح.، (۱۳۹۳): ترمیم و گسترش ۱۲۷ سال دمای ماهانه مشهد، *مجله پژوهش های اقلیم شناسی*، ۵ (۱۷ و ۱۸): ۱۲۳-۱۱۱.

۷. مطیع قادر، ح.، لطفی، ش.، سیداسفهلان، م.م، (۱۳۸۹): *مروری بر برخی روش های بهینه سازی هوشمند*، چاپ دانشگاه آزاد اسلامی واحد شبستر.

8. Abbass H. A., Sarker R. A., and Newton C. S., 2002. Data Mining, a Heuristic Approach-IGI Global. Idea Group Publishing. pp300.
9. Aydilek I. B., Arslan A., 2013. A hybrid method for imputation of missing values using optimized fuzzy c-means with support vector regression. Information Sciences, 233 (1): 25–35.
10. Chaudhuri, S., Goswami, S., Das, D., Middey, A., 2014. Meta-heuristic ant colony optimization technique to forecast the amount of summer monsoon rainfall: skill comparison with Markov chain model. Theoretical and Applied Climatology,116 (3–4):585–595.
11. Dastorani, M. T., Moghadamnia, A., Piri, J., Rico-Ramirez, M., 2010, "Application of ANN and ANFIS models for reconstructing missing flow data." Environmental monitoring and assessment 166.,1-4: 421-434.
12. Dingman, S. L., 2002, Physical Hydrology, Second Edition, PRENTICE HALL.
13. Dipak, V. P., Bichkar, R. S., (2010): Multiple Imputation of Missing Data with Genetic Algorithm based Techniques, Evolutionary Computation for Optimization Techniques.
14. Ghahraman, B., Ahmadi, F., (2007): Application of Geo statistics in Time series: Mashhad Annual Rainfall, Iran-Watershed Management Science & Engineering. Vol. 1, No. 1.
15. Jacob. D., Reed. D. W., Robson. A. J., (1999): Choosing a pooling group. Flood Estimation Handbook. Vol. 3. Institute of Hydrology, Wallingford, UK.
16. Ranhao, S., Baiping, Z., and Jing, T., 2008, A Multivariate Regression Model for Predicting Precipitation in the Daqing Mountains, Mountain Research and Development, 28(3):318-325.
17. Smithsonian Institution. (1927, 1934, 1947): World weather records, 1910-1920., 1921-1930., 1931 – 1940., Smithson. Miss C. Collect. 79,90,105. (Publication2913.,3216.,3803)
18. Ustoorikar, K., Deo, M.C., (2008): Filling up gaps in wave data with genetic programming, Marine Structures, 21, 177–195.
19. Yozgatligil, C., Aslan S., Iyigun, C., Batmaz, I., (2013): Comparison of missing value imputation methods in time series: the case of Turkish meteorological data, Theory Apply Climatology, 112:143–167.
20. Little, R. JA, Rubin. D., B., 2002, Statistical analysis with missing data. John Wiley & Sons, 408 pages.
21. Patil, D V, Bichkar, R S. 2010, Multiple Imputation of Missing Data with Genetic

Algorithm based Techniques. IJCA Special Issue on "Evolutionary Computation for Optimization Techniques.

22. El Assaad H., Samé, A., Govaert, G., Aknin, P., 2016, A variational Expectation–Maximization algorithm for temporal data clustering, Computational Statistics and Data Analysis,103:206–228.
23. Preis,A., Ostfeld, A., 2008, A coupled model tree–genetic algorithm scheme for flow and water quality predictions in watersheds, Journal of Hydrology (Elsevier), 349: 364–375.
24. Iqbal M, WEN J, WANG Sh, TIAN Hu, ADNAN M. 2018, Variations of precipitation haracteristics during the period 1960-2014 in the Source Region of the Yellow River, China. Journal of Arid Land, 10(3): 388-401.