

## بررسی کارآیی دو روش داده محور در پیش‌بینی بارندگی ماهانه

مسلم برجی حسن گاویار<sup>۱</sup>، علیرضا مقدم نیا<sup>۲\*</sup>، فرزانه ساجدی<sup>۳</sup>

۱- کارشناسی ارشد مهندسی آبخیزداری، دانشگاه تهران

۲- دانشیار گروه احیاء مناطق خشک و کوهستانی دانشکده منابع طبیعی، دانشگاه تهران

۳- کارشناسی ارشد مهندسی آبخیزداری، دانشگاه کشاورزی و منابع طبیعی ساری

پذیرش مقاله: ۱۳۹۵/۴/۱۲

تأیید نهایی مقاله: ۱۳۹۶/۴/۱۵

### چکیده

پیش‌بینی دقیق بارش باران به خصوص در مناطق خشک و نیمه خشک، نقش مهمی را در مدیریت موثر منابع آب بازی می‌کند. حصول روش‌های مناسب و دقیق در پیش‌بینی بارندگی یکی از مسائل چالش انگیز امروزی، در مدیریت منابع آب و مخاطرات اقلیمی است. اگرچه تحقیقات گسترده‌ای در زمینه کاربرد مدل‌های هوش محاسباتی به منظور پیش‌بینی‌های اقلیمی صورت گرفته است، اما انتخاب نوع و تعداد متغیرهای ورودی به هر کدام از این مدل‌ها همواره مدل‌سازان را با مسائلی روبرو کرده است. هدف از این تحقیق، بررسی تأثیر پیش‌پردازش داده‌ها در انتخاب بهترین ترکیب ورودی از متغیرهای تأثیرگذار بر فرآیند بارش با استفاده از آزمون گاما برای پیش‌بینی بارش ماهانه با دو مدل رگرسیون بردار پشتیبان و برنامه‌ریزی بیان ژن می‌باشد. برای این منظور، از روش‌های آزمون گاما و آنالیز همبستگی برای پیش‌پردازش ورودی مدل‌های مورد استفاده در این تحقیق تحت یک مطالعه موردی با استفاده از داده‌های اقلیمی ماهانه مربوط به ایستگاه سینوپتیک شیراز در طی سال‌های ۱۳۶۲ تا ۱۳۹۰ استفاده شد. کارآیی این مدل‌ها با استفاده از ضرایب تبیین، ریشه میانگین مربعات خطا و ضریب کارآیی ناش-ساتکلیف ارزیابی گردید. نتایج نشان داد که مدل ترکیبی گاما-رگرسیون بردار پشتیبان، بارندگی ماهانه را بهتر از سایر مدل‌های استفاده شده در این تحقیق پیش‌بینی می‌کند. ولی آزمون گاما نتوانست کارایی مدل برنامه‌ریزی بیان ژن را به اندازه مدل رگرسیون بردار پشتیبان بهبود بخشد. همچنین براساس نتایج حاصله، متغیرهای ساعات آفتابی، رطوبت نسبی، بارندگی یک ماه گذشته و دما به ترتیب جزء موثرترین متغیرها در پیش‌بینی بارندگی ماهانه می‌باشند.

**واژه‌های کلیدی:** آزمون گاما، بارندگی، برنامه‌ریزی بیان ژن، پیش‌بینی، رگرسیون بردار پشتیبان.

## مقدمه

(۱۳۸۹) فرآیند بارش-رواناب حوزه آبریز لیقوان با استفاده از برنامه‌ریزی ژنتیک، ابتدا با ۱۵ متغیر برای تعیین متغیرهای معنی‌دار و سپس مدل‌سازی نهایی با متغیرهای معنی‌دار و دو مجموعه از عملگرهای ریاضی را مدل‌سازی کردند و دریافتند که برنامه‌ریزی ژنتیک می‌تواند در زمینه پیش‌بینی جریان و نیز مدل‌سازی بارش-رواناب کاربرد مناسبی داشته باشد. رابونال و همکاران (۲۰۰۷) با استفاده از برنامه‌ریزی ژنتیک و شبکه عصبی مصنوعی نسبت به ارائه آبنمود واحد در حوزه‌های شهری اقدام نمودند. آنها از GP برای پیش‌بینی بارندگی و اثر آن بر رواناب، و از ANN برای پیش‌بینی اثرات انسانی بر توالی رواناب استفاده کرده‌اند. مطالعات زیادی در زمینه کارآیی بهتر روش برنامه‌ریزی بیان ژن نسبت به مدل شبکه عصبی مصنوعی انجام شده است. با مروری بر تحقیقات سال‌های اخیر در زمینه پیش‌بینی بارش این نتیجه حاصل شد که بیشتر محققان تا سال ۲۰۱۰ از روش ANN و یا به صورت ترکیبی با مدل‌های خطی استفاده نمودند (کرامز و همکاران، ۲۰۰۹؛ ورسینگ و همکاران، ۲۰۱۰؛ ویو و همکاران، ۲۰۱۰؛ پاتیل و همکاران، ۲۰۱۰). از سال ۲۰۱۰ به بعد رویکرد محققان به سمت استفاده از روش‌های مانند تحلیل موجک، ماشین بردار پشتیبان، برنامه‌ریزی بیان ژن تغییر یافته است. علاوه بر نوع روش به کار برده شده برای پیش‌بینی، تعیین مناسب‌ترین متغیرهای ورودی مدل نیز یکی از موارد مهم پیش‌بینی می‌باشد (قبایی سوغ و همکاران، ۲۰۱۰). در تحقیقات گذشته تأثیر پارامترهای اقلیمی مختلف بر پیش‌بینی بارندگی همواره مورد توجه قرار گرفته است. احمدی و همکاران (۱۳۹۳) بارش دراز مدت حوزه آبریز اهرچای را بر پایه الگوهای پیوند دور

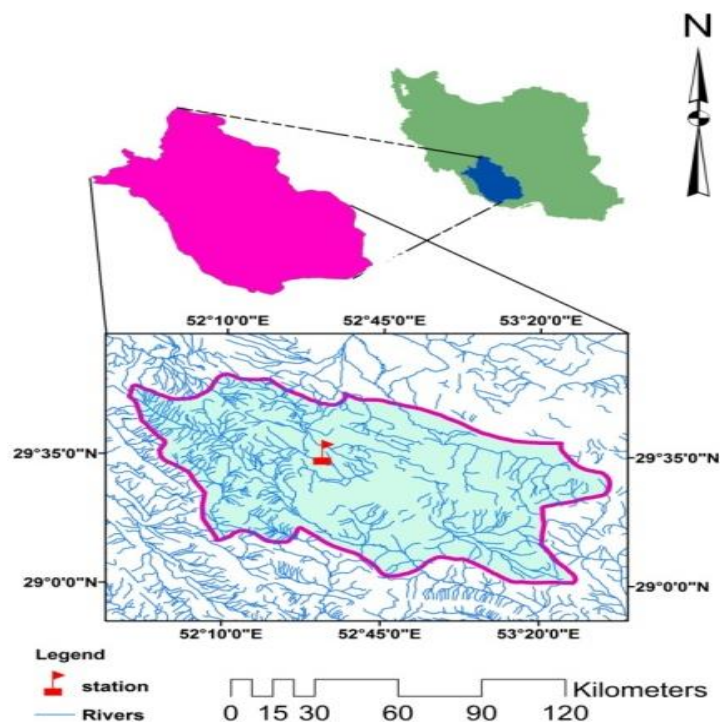
اطلاعات دقیق از بارندگی امری حیاتی در مدیریت مخازن سدها و کنترل سیل است. همچنین بارندگی تأثیر زیادی بر روی ترافیک، سامانه زهکش فاضلاب و دیگر فعالیت‌های بشری در جوامع شهری دارد. بارندگی یکی از پیچیده‌ترین و اتفاقی‌ترین پدیده‌های طبیعی است. در واقع، تأثیر عوامل محسوس و یا غیرمحسوس آن‌چنان در ایجاد حادثه دخیل است که این روند را از یک نظام قانونمند روشن به سوی ساختاری پیچیده و آشوبناک سوق داده است (گلابی و همکاران، ۱۳۹۳). در دهه اخیر تحقیقات گسترده‌ای در زمینه پیش‌بینی بارندگی (ماهانه و فصلی) با استفاده از سیگنال‌های بزرگ مقیاس اقلیمی (احمدی و همکاران، ۲۰۱۴) و متغیرهای محلی (دما، رطوبت، فشار، تبخیر و...) انجام شده است (لین و همکاران، ۲۰۱۳؛ چان، ۱۹۹۷) مدل‌های خطی و غیرخطی متعددی بدین منظور بکار گرفته شده است، که نوع غیرخطی آنها مانند شبکه عصبی مصنوعی، ماشین بردار پشتیبان و برنامه‌ریزی بیان ژن از توانایی بالاتری برخوردار بوده‌اند (نجفی و همکاران، ۲۰۱۰؛ ال-شفیع و همکاران، ۲۰۱۱). در دهه اخیر ماشین بردار پشتیبان (SVM) در زمینه‌های متعدد هیدرولوژیکی بکار گرفته شده است. گیو در سال ۲۰۱۱ با استفاده از روش ANN، SVR، LSSVR، LR، RR مقادیر بارش را در اوکلاهما تخمین زده‌اند. آنها در نهایت روش رگرسیون بردار پشتیبان را به‌عنوان روش مناسب معرفی کردند. سیواپراگاسامت و همکاران (۲۰۰۵) از مدل ترکیبی SVM و SSA برای پیش‌بینی بارش و رواناب استفاده نمودند. نتایج نشان داد که مدل ترکیبی در مقایسه با مدل اصل SVM دارای عملکرد بهتری بوده است. سلطانی و همکاران

پیش‌بینی بارندگی و انتخاب ترکیب ورودی مناسب از دو روش آزمون گاما و آزمون همبستگی استفاده کردیم.

#### محدوده مورد مطالعه

استان فارس دارای متوسط بارندگی حدود ۳۳۵ میلیمتر و وسعت ۱۲۲۶۰۹ کیلومتر در ارتفاع ۱۴۹۱ متری از سطح دریا قرار دارد که شیب کلی دشت شیراز از همه جهات به طرف دریاچه مهارلو به شدت کاهش می‌یابد و در جنوب ایران قرار دارد. شهرستان شیراز از بزرگترین شهر حوزه جنوبی ایران است که با وسعت ۱۰۴۳۴ کیلومتر مربع ۸/۵ از کل مساحت خاکی استان فارس را به خود اختصاص داده است. این شهرستان در محدوده جغرافیایی ۵۳ درجه و ۳۷ دقیقه طول شرقی و ۲۹ درجه و ۵۷ دقیقه عرض شمالی قرار گرفته است (شکل ۱). متوسط بارش این شهرستان حدود ۳۳۵ میلیمتر و متوسط دما سالانه آن نیز حدود ۱۸ درجه سانتی‌گراد است. در این تحقیق از داده‌های بارندگی ماهانه، دما میانگین، تبخیر، سرعت باد، ساعات آفتابی و رطوبت نسبی ایستگاه سینوپتیک شیراز در طی سال‌های ۱۳۶۲ تا ۱۳۹۰ استفاده شده است.

اقلیمی، با استفاده از مدل SVR پیش‌بینی کردند، و از روش‌های همبستگی و آزمون گاما برای انتخاب بهترین ترکیب ورودی مدل استفاده شده است، و نتایج نشان داده است که انتخاب ورودی‌های مدل با روش آزمون گاما خطای شبیه‌سازی کمتری نسبت به روش رگرسیون در برآورد میزان بارندگی دارد. دستورانی و همکاران (۱۳۹۱) به منظور مدل‌سازی بارش-رواناب در حوزه آبخیز سد زاینده رود با استفاده از شبکه عصبی مصنوعی و سامانه استنتاج عصبی-فازی تطبیقی، از الگوریتم ژنتیک و آزمون گاما برای تعیین بهترین ترکیب ورودی از متغیرها استفاده کرده‌اند. با مرور انجام گرفته بر منابع گذشته، ضعف موجود در تعیین متغیرهای مناسب برای پیش‌بینی بارندگی توسط مدل‌های مختلف و کمبود روش سیستماتیک برای انتخاب بهترین ترکیب از میان متغیرهای موجود به عنوان ورودی به این مدل‌ها مشخص شد. به همین دلیل، در این تحقیق یک مدل پیش‌بینی بارندگی با ترکیب بهینه از متغیرهای ورودی ارائه شده است. کارایی دو مدل داده‌محور ماشین بردار پشتیبان و برنامه‌ریزی بیان ژن برای پیش‌بینی ماهانه بارش ایستگاه سینوپتیک شیراز بررسی شده است. همچنین برای تعیین متغیرهای اقلیمی موثر بر



شکل ۱: منطقه مورد مطالعه و موقعیت ایستگاه سینوپتیک شیراز را نشان می‌دهد.

## مواد و روش‌ها

### آزمون گاما

انتخاب متغیرهای ورودی برای مدل‌هایی مانند رگرسیون چند متغیره و مدل‌های غیرخطی مانند ANN یک مشکل پیچیده و از مراحل مهم مدل‌سازی می‌باشد (چانگ و همکاران، ۲۰۱۵). در این تحقیق از مدل جدید آزمون گاما و همبستگی برای شناسایی بهترین ترکیب ورودی به مدل‌های SVR و GEP استفاده شده است. آزمون گاما یک مدل‌سازی غیرخطی و ابزاری جهت تحلیل می‌باشد و اجازه می‌دهد تا رابطه بین ورودی‌ها و خروجی‌ها در یک مجموعه داده‌های عددی مورد امتحان قرار گیرد. آزمون گاما حداقل متوسط مجذور خطایی مدل را پیش از استفاده محاسبه و ترکیب بهینه متغیرهای ورودی شناسایی می‌شوند (مقدم نیا و همکاران، ۲۰۰۹). در واقع آزمون گاما تخمینی از حداقل میانگین مربعات خطای آن بخش از داده‌ها است که با مدل خطی نمی‌توان آنها را تعیین کرد.

فرض کنید یک سری داده مشاهداتی به صورت رابطه ۱ وجود دارد:

رابطه ۱)

$$(X_1, \dots, X_m, y) = (X, y)$$

که در آن  $X = (X_1, \dots, X_m)$  بردار ورودی (مثلاً ساعات آفتابی، رطوبت و...) در محدوده است و  $y$  بردار خروجی (بارندگی ماهانه ایستگاه شیراز). اگر رابطه ۲ بین ورودی‌ها و خروجی‌ها برقرار باشد:

رابطه ۲)

$$y = f(x_1, \dots, x_m) + r$$

در آن  $r$  یک متغیر تصادفی است که نشان دهنده خطا است. آزمون گاما تخمینی برای واریانس خروجی مدلی غیرهموار است. آماره گاما ( $\Gamma$ ) واریانس بخشی از داده‌ها را که تابع  $f$  قادر به توجیه آن نیست نشان می‌دهد. آزمون گاما براساس است که شامل لیستی از امین همسایگی برای هر بردار است. تابع دلتا میانگین مجذور فاصله  $K$  امین

باد (WV)، ساعات آفتابی (Sh)، رطوبت نسبی (H) و تأخیر یک ماه پیشین از متغیرهای نامبرده (-T- 1، E-1، WV-1، Sh-1 و H-1) و همچنین بارندگی یک ماه پیش و بارندگی ماهانه به عنوان خروجی متناظر آنها در نظر گرفته شد. دلیل عدم انتخاب تأخیرهای بیش از یک ماه در هر کدام از متغیرهای ورودی، عدم همبستگی معنی دار با مقدار متغیر هدف (بارندگی) بوده است. بعضی اوقات انتخاب متغیرهای زیاد به عنوان ورودی موجب پیچیدگی مدل خواهد شد و یا اینکه از متغیرهایی که تأثیر ملموسی بر نتایج پیش بینی نمی گذارند استفاده می شود. در این تحقیق از آزمون گاما برای انتخاب موثرترین متغیرها استفاده شد. بدین منظور با استفاده از نرم افزار WinGama ابتدا تمام متغیرها انتخاب گردیدند و آزمون گاما بر روی آنها اعمال شد. سپس به ترتیب هر کدام از متغیرها یکبار از ترکیب حذف گردیده و نتایج براساس آماره گاما متناظر با هر متغیر حذف شده بررسی شد. اگر میزان آماره گاما، با حذف یک متغیر افزایش (نسبت به مقدار آن در حالتی که کل ورودیها انتخاب شده اند) یابد، یعنی وجود آن متغیر تأثیر معنی داری بر نتایج پیش بینی خواهد گذاشت.

#### رگرسیون بردار پشتیبان

روش ماشین های بردار پشتیبان، دسته ای از روش های یادگیری با ناظر هستند که برای مسائل طبقه بندی و رگرسیون به کار می روند. در ابتدا SVM به طور معمول برای حل مسئله طبقه بندی مورد استفاده قرار گرفت، سپس نسخه دیگری از SVM برای مشکل رگرسیون توسط وپنک (۲۰۱۳) پیشنهاد شد. نسخه جدید، رگرسیون بردار پشتیبان ۸ نامیده می شود که رایج ترین شکل مورد استفاده از ماشین های بردار پشتیبان است. اگر نقاط آموزشی را به صورت، بردار ورودی را و ارزش طبقه را تعریف کنیم، آنگاه در حالتی که داده ها به صورت

همسایگی را محاسبه می کند (احمدی و همکاران، ۲۰۰۹).

رابطه ۳)

$$\delta_m(k) = \frac{1}{M} \sum_{i=1}^M |X_{N[i,k]} - X_i|^2$$

که در آن نشان دهنده فاصله اقلیدسی است، تابع گاما متناظر آن:

رابطه ۴)

$$\gamma_M(K) = \frac{1}{2M} \sum_{i=1}^M (y_{N[i,k]} - y_i)^2$$

که در آن،  $m(k)Y$  بیانگر مقادیر  $y$  برای  $k$  امین همسایه  $x_i$  در رابطه ۳ است. از این رو با محاسبه  $\Gamma$  می توان کمترین مربعات خطا را برای نقاط  $[\delta_m p$  (k),  $Y_M$  (k)] ایجاد کرد:

رابطه ۵)

$$\gamma = A\delta + \Gamma$$

عرض از مبدأ این خط، مقدار را نشان می دهد و برابر با واریانس خطاهاست. اگر  $m$  تعداد متغیرهای ورودی باشد، در این صورت ترکیب از میان آنها وجود خواهد داشت. بررسی همه این ترکیبات به زمان زیادی نیاز دارد. آزمون گاما می تواند موثرترین متغیرها در مدل سازی و بهترین ترکیب از متغیرهای ورودی را مشخص کند. همچنین برای تعیین طول دوره آموزش از روش آزمون  $M$  و براساس آماره گاما استفاده می شود (قبایی سوغ و همکاران، ۲۰۱۰). آزمون  $M$  در واقع رسم نمودار تغییرات مقدار آماره  $\Gamma$  و یا خطای استاندارد در مقابل تعداد داده ها می باشد. در نقطه ای که این دو نمودار تقریباً به حالت افقی و پایدار می رسند آن نقطه تعیین کننده تعداد نقاط کافی برای آموزش مدل خواهد بود (شریفی و همکاران، ۱۳۹۳). در این تحقیق بارندگی ماهانه ایستگاه سینوپتیک شیراز با استفاده از مدل های SVR و GEP پیش بینی شد. داده های ورودی به مدل شامل مقادیر ماهانه دما میانگین (T)، تبخیر (E)، سرعت

با تابع کرنل RBF استفاده گردید. محاسبات مربوط به رگرسیون بردار پشتیبان در این تحقیق با استفاده از نرم‌افزار R انجام گرفته است. به منظور آموزش کرنل، براساس آزمون M و مشاهده ضرایب کارایی از ۸۰٪ داده‌های موجود، استفاده گردید و مابقی داده‌ها برای آزمون کرنل بکار گرفته شد.

#### برنامه‌ریزی بیان ژن (GEP)

برنامه‌ریزی بیان ژن توسط فریرا در سال ۱۹۹۹ ابداع شد. در این برنامه، کروموزوم‌های خطی و ساده با طول ثابت، مشابه با آنچه که در الگوریتم ژنتیک استفاده می‌شود و ساختارهای شاخه‌ای با اندازه‌ها و اشکال متفاوت، مشابه با درختان تجزیه در برنامه‌ریزی ژنتیک، ترکیب می‌شوند. از آنجائی- که تمامی ساختارهای شاخه‌ای با اندازه‌ها و اشکال متفاوت، در کروموزوم‌های خطی با طول ثابت کدگذاری می‌شوند، معادل این است که بگوییم در GEP، ژنوتیپ و فنوتیپ سرانجام از یکدیگر جدا شده و اکنون سیستم می‌تواند از تمام مزایای تکاملی به سبب وجود آنها بهره‌برد. با وجود اینکه فنوتیپ در GEP، همان نوع از ساختارهای شاخه‌ای مورد استفاده در GP را شامل می‌شود، اما ساختارهای شاخه‌ای که بوسیله GEP استنتاج می‌شوند (که بیان درختی نیز نامیده می‌شوند) مبین تمامی ژنوم‌های مستقل هستند (فریرا، ۲۰۰۶). مراحل اصلی از برنامه‌ریزی بیان ژن به طور شماتیک در شکل ۲ نشان داده شده است. همان-طور که مشاهده می‌شود، فرآیند با تولید تصادفی کروموزوم‌ها از تعداد معینی افراد (جمعیت اولیه) آغاز می‌شود. سپس این کروموزوم‌ها به صورت بیان درختی (ET) اظهار می‌شوند و میزان برازندگی هر فرد در مقابل مجموعه‌ای از موارد برازش، ارزیابی می‌گردد (که انتخاب محیط نیز نامیده می‌شود). افراد سپس مطابق شایستگی خود (عملکرد آنها در آن محیط ویژه) انتخاب می‌شوند تا با اصلاحات و

خطی قابل تفکیک هستند، قواعد تصمیم‌گیری تعریف می‌شود و با یک صفحه بهینه که طبقات تصمیم‌گیری دودویی را تفکیک می‌کند، به صورت رابطه ۶ خواهد بود.

رابطه ۶)

$$y = \text{sign} \left( \sum_{i=1}^N y_i \alpha_i (X \cdot X_i) + b \right)$$

که در آن  $y$  خروجی معادله،  $y_i$  ارزش طبقه نمونه آموزشی و نشان دهنده ضرب داخلی است. بردار  $x=(x_1, x_2, \dots, x_n)$  نشان دهنده یک داده ورودی و بردارهای  $X_i$  ( $i=1, \dots, N$ ) بردارهای پشتیبان هستند. در رابطه ۶ پارامترهای  $b$  و  $a_i$  تعیین کننده ابر صفحه هستند. اگر داده‌ها به صورت خطی قابل تفکیک نباشند، رابطه ۶ به صورت رابطه ۷ تغییر می‌یابد:

رابطه ۷)

$$y = \text{sign} \left( \sum_{i=1}^N y_i \alpha_i K(X \cdot X_i) + b \right)$$

تابع  $K(X, X_i)$  تابع کرنلی است که برای ایجاد ماشین‌هایی با انواع مختلفی از سطوح تصمیم‌گیری غیرخطی در فضای داده‌ها، ضرب‌های داخلی تولید می‌کند. برای مثال سه نوع تابع کرنل که در مدل SVM به کار می‌روند، عبارتند از: تابع کرنل چندجمله‌ای، تابع کرنل پایه شعاعی و تابع کرنل پرسپترون چندلایه. قابلیت به کارگیری کرنل‌ها یکی از نقاط قوت کلیدی SVMها محسوب شده، که سبب می‌شود آنها انعطاف بالایی برای حل انواع مختلف مسائل با دشواری‌های گوناگون داشته باشند (کریستیانین و همکاران، ۲۰۰۰) دو نوع مدل رگرسیون بردار پشتیبان تعریف شده است که نوع اول SVM- $\nu$  و نوع دوم SVM- $\epsilon$  نامیده می‌شوند. تحقیقات اخیر نشان داده که مدل SVM- $\nu$  با تابع کرنل پایه شعاعی دارای نتایج بهتری نسبت به نوع دوم است (احمدی و همکاران، ۲۰۱۴؛ نوری و همکاران، ۲۰۰۹) بدین خاطر در این تحقیق از مدل

خام باعث کاهش سرعت و دقت شبکه می‌شود. برای احراز از چنین شرایطی و همچنین به منظور یکسان کردن ارزش داده‌ها برای شبکه، عمل نرمال‌سازی صورت گرفت. این کار مانع از کوچک شدن بیش از حد وزن‌ها و سبب جلوگیری از اشباع زود هنگام نرون‌ها می‌شود. غالباً، نرمال کردن داده‌ها در بازه ۰/۱ و ۰/۹ صورت می‌گیرد. در این تحقیق کلیه داده‌های ورودی قبل از اعمال به مدل GEP، با استفاده از رابطه ۸ نرمال گردیدند:

رابطه ۸)

$$X_{normal} = 0.1 + 0.8 \times \left( \frac{X_o - X_{min}}{X_{max} - X_{min}} \right)$$

#### ارزیابی کارایی مدل‌های مورد استفاده

نتایج مدل‌های بکار برده شده در این تحقیق به وسیله ریشه میانگین مربعات خطا، ضریب همبستگی و ضریب کارایی ناش-ساتکلیف بین مقادیر مشاهده‌ای و پیش‌بینی شده مورد ارزیابی قرار گرفته است.

ضریب همبستگی: (R)

رابطه ۹)

$$R = \frac{\sum_{i=1}^n (X_o - \bar{X}_o)(X_e - \bar{X}_e)}{\sqrt{\sum_{i=1}^n (X_o - \bar{X}_o)^2 \sum_{i=1}^n (X_e - \bar{X}_e)^2}}$$

ضریب ناش و ساتکلیف:

رابطه ۱۲)

$$NS = 1 - \frac{\sum_{i=1}^n (X_o - X_e)^2}{\sum_{i=1}^n (X_o - \bar{X}_o)^2}$$

که در آنها SEE میانگین مربعات خطاست و N تعداد نمونه‌های استفاده شده، X<sub>0</sub> مقدار مشاهده

بهبودهایی دوباره ساخته شوند و در نتیجه فرزندان با ویژگی‌های جدید باقی می‌گذارند. این فرزندان جدید، در یک چرخه، تحت همان فرآیند توسعه‌ای قرار می‌گیرند یعنی بیان ژنوم، مواجهه با انتخاب محیط، انتخاب بر مبنای برآزش، و تولید مثل با بهبود. این فرآیند برای تعداد معینی از نسل‌ها تکرار می‌شود تا راه‌حل خوب و مناسبی پیدا شود (لوپس و وینرت، ۲۰۱۴). قابل ذکر است که تولید مثل تنها شامل تکثیر نیست بلکه، فعالیت عملگرهای ژنتیکی توانا در ایجاد تنوع ژنتیکی را نیز در بر می‌گیرد. در واقع تولید مثل، شامل تکثیر و اصلاحات ژنتیکی است. تکثیر عملیاتی است که چندین فرد مناسب از نسل حاضر را برای نسل بعد نگه می‌دارد. در طول تکثیر، ژنوم کپی شده به نسل بعد انتقال داده می‌شود. بدیهی است، تکثیر به تنهایی نمی‌تواند تغییرات را نشان دهد و فقط به واسطه فعالیت بقیه عملگرها، تغییرات ژنتیکی داخل جمعیت نشان داده می‌شود. این عملگرها، کروموزوم‌ها را به طور تصادفی انتخاب می‌کنند تا اصلاح شوند. بنابراین در GEP یک کروموزوم در یک زمان ممکن است بوسیله یک یا چند عملگر ژنتیکی اصلاح شود و یا اصلاً مورد اصلاح قرار نگیرد (قربانی و همکاران، ۲۰۱۰). در این پژوهش برای به کارگیری مدل برنامه‌ریزی بیان ژن از نرم افزار GeneXpro4.3 استفاده گردید. اصولاً، وارد کردن داده‌ها به صورت

ریشه میانگین مربعات خطا: (RMSE)

رابطه ۱۰)

$$RMSE = \sqrt{\frac{SSE}{N}}$$

که در آن است:

رابطه ۱۱)

$$SSE = \sum_{i=1}^n (X_o - X_e)^2$$

شده بارندگی،  $X_e$  مقدار برآورد شده بارندگی و  $\bar{X}_e$  و  $\bar{X}_o$  میانگین داده‌های مشاهداتی و تخمین زده شده هستند (شکل ۲).



شکل ۲: فلوچارت مراحل تحقیق

### نتایج

آموزش مدل‌ها تأثیری بر کارآیی آنها نخواهد گذاشت. صحت سنجی این آزمون از طریق انتخاب تعداد متفاوتی از داده‌ها برای آموزش مدل‌ها و سپس بررسی ضرایب کارآیی صورت گرفت و در نهایت ۸۰ درصد داده برای آموزش مدل استفاده شد. البته ذکر این نکته الزامی است که انتخاب دامنه اعداد برای آموزش، بین ۷۰ تا ۹۰ درصد از داده‌ها صورت گرفت و در دامنه  $5\pm/80$  تعداد داده‌ها، تفاوت ملموسی در نتایج ایجاد نکردید؛ و

نتایج مربوط به حذف هر کدام از متغیرهای ورودی در جدول ۱ آورده شده است. همچنین برای انتخاب طول دوره آماری مناسب از روش آزمون M استفاده شد. شکل ۳ منحنی مقادیر آماره گاما را در مقابل تعداد داده‌ها نشان می‌دهد. همان‌طور که مشخص است از داده ۱۲۶۸م (۸۰ درصد داده‌های موجود) به بعد تقریباً نوسان مقادیر گاما کمتر شده است. این بدین معنی است که انتخاب تعداد داده‌های بیشتر یا کمتر از این مقدار برای



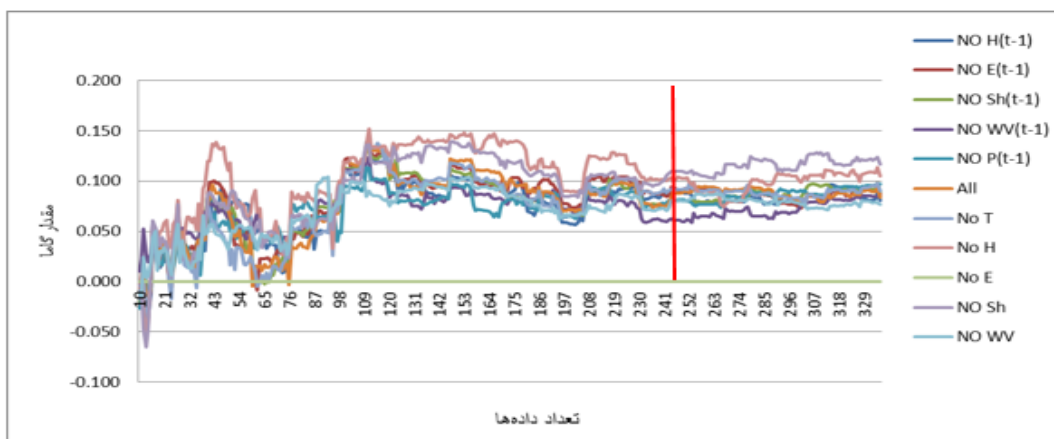
به این دلیل ۸۰٪ داده‌ها برای آموزش و ۲۰ درصد باقی مانده برای آزمایش مدل به کار گرفته شد.

جدول ۱: نتایج آزمون گاما در انتخاب بهترین پیش‌بینی کننده‌ها

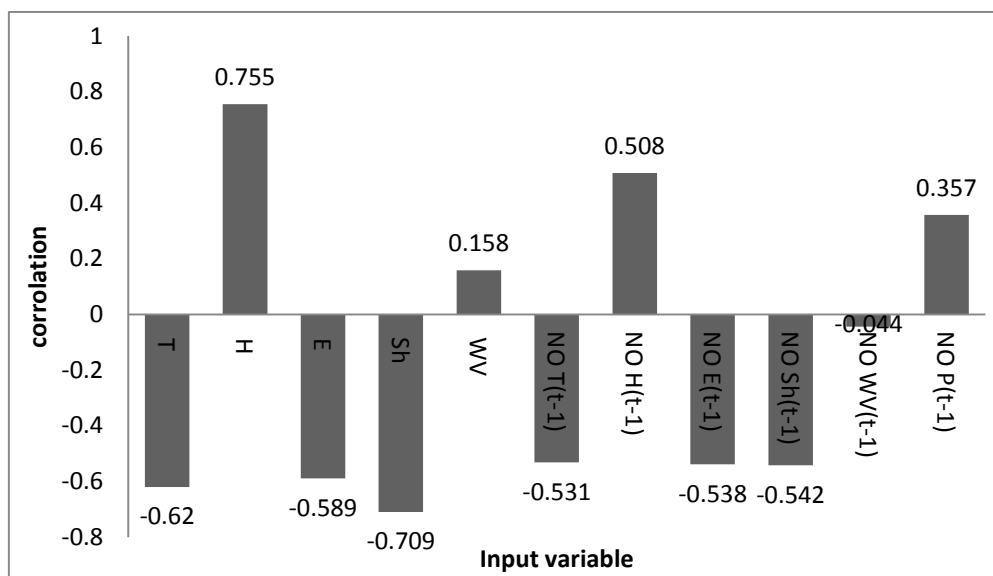
ترکیب ورودی‌ها	گاما	خطای استاندارد	ترکیب ورودی‌ها
<i>All</i>	۰/۰۵	۰/۰۳۳	۱۱۱۱۱۱۱۱۱۱
<i>No T</i>	۰/۰۹	۰/۰۰۸	۰۱۱۱۱۱۱۱۱۱۱
<i>No H</i>	۰/۱۰۵۲	۰/۰۱۲۲	۱۰۱۱۱۱۱۱۱۱۱
<i>No E</i>	۰/۰۸۹۶	۰/۰۰۷	۱۱۰۱۱۱۱۱۱۱۱
<i>No Sh</i>	۰/۱۱۶۸	۰/۰۱۴۵	۱۱۱۰۱۱۱۱۱۱۱
<i>No WV</i>	۰/۰۷۶۶	۰/۰۱۱۲	۱۱۱۱۰۱۱۱۱۱۱
<i>No T-I</i>	۰/۰۸۰۹	۰/۰۰۹	۱۱۱۱۱۰۱۱۱۱۱
<i>No H-I</i>	۰/۰۸۰۵	۰/۰۰۸	۱۱۱۱۱۱۰۱۱۱۱
<i>No E-I</i>	۰/۰۸۶۵	۰/۰۰۷۴	۱۱۱۱۱۱۱۰۱۱۱
<i>No Sh-I</i>	۰/۰۸۹۵	۰/۰۰۵۵	۱۱۱۱۱۱۱۱۰۱۱
<i>No Wv-I</i>	۰/۰۸۴۶	۰/۰۱۱۱	۱۱۱۱۱۱۱۱۱۱۰۱
<i>No P-I</i>	۰/۰۹۶۸	۰/۰۱۶۷	۱۱۱۱۱۱۱۱۱۱۰

کرده‌است. پس وجود این دو متغیر در ترکیب ورودی برای مدل‌های پیش‌بینی لازم است. سایر متغیرهای تأثیرگذار به ترتیب شامل بارندگی یک ماه گذشته (P-1) و دما (T) هستند.

نتایج جدول ۱ نشان می‌دهد که متغیر ساعات آفتابی (Sh) و متوسط رطوبت نسبی (H) بیشترین تأثیر را در پیش‌بینی بارندگی ماهانه دارند. آماره گاما با حذف هر کدام از این دو متغیر نسبت به حالتی که کل متغیرها وجود دارند افزایش پیدا



شکل ۳: تغییر مقدار گاما در مقابل تعداد داده‌ها



شکل ۴: میزان همبستگی بین متغیرهای ورودی و بارندگی

بودند به‌عنوان ترکیب‌های ورودی‌ها انتخاب شدند. براساس فرضیه این تحقیق، سناریوی اول که در آن متغیرهای H، E، Sh، Sh-1 و Wv-1 وجود دارند، با توجه به آماره گامای پایین انتظار می‌رود که در صورت انتخاب به‌عنوان ورودی بهترین عملکرد را برای مدل‌های SVR و GEP داشته باشد. سناریوهای دو، سه و چهار نیز به ترتیب براساس آماره گاما کمتر به بیشتر برای مقایسه با سناریوهای دیگر در جدول یک نشان داده شده است. سناریو پنجم نیز که براساس میزان ضریب همبستگی بین متغیرها ورودی با متغیر هدف انتخاب گردیده است، عوامل T، H، E و Sh را به‌عنوان موثرترین متغیرها معرفی می‌کند. سناریوی ششم که کلیه متغیرها را برای ورود به مدل‌ها در نظر گرفته است، به‌عنوان سناریوی شاهد معرفی می‌شود و نتایج سایر سناریوها با آن مقایسه خواهد شد.

سایر متغیرها دارای تأثیر یکسانی بر پیش‌بینی هستند. شکل ۴ مقادیر همبستگی هرکدام از متغیرهای انتخابی با میزان بارندگی ثبت شده در ایستگاه سینوپتیک شیراز را نشان می‌دهد. بیشترین همبستگی را متغیرهای رطوبت (H)، ساعات آفتابی (Sh) و دما (T) دارد که با سناریوهای معرفی شده به وسیله آزمون گاما مطابقت نسبی دارد. در مجموع، انتخاب بهترین ترکیب از این متغیرها، با استفاده از آماره گاما و آزمون همبستگی بین متغیرهای ورودی با مقادیر بارندگی متناظر، مشخص گردید، که در جدول ۲ در قالب پنج سناریو تعریف شده‌اند. با توجه به ۱۱ متغیر ورودی،  $2047 - 1 = 211$  ترکیب ورودی قابل تشکیل است، که ارزیابی تمامی آنها به‌عنوان ترکیب ورودی مدل نیازمند وقت زیادی است. پس چهار ترکیب که کمترین میزان آماره گاما و همچنین یک ترکیب شامل متغیرهایی که همبستگی آنها با متغیر هدف بیشتر از ۵۵ درصد

جدول ۲: نتایج آزمون گاما و همبستگی برای انتخاب بهترین ورودی مدل

		گاما	ترکیب ورودی
سناریوهای روش گاما	سناریو اول	۰/۰۳۷۷	۰۱۱۱۰۰۰۰۱۱۰
	سناریو دوم	۰/۰۳۸۸	۱۱۱۱۰۰۱۰۱۰۰
	سناریو سوم	۰/۰۳۹۷	۰۱۱۱۰۰۱۰۱۱۰
	سناریو چهارم	۰/۰۴۰۴	۰۱۰۱۰۰۱۱۱۱۰
سناریوی روش همبستگی	سناریو پنجم	-	۱۱۱۱۰۰۰۰۰۰۰
سناریو شاهد	سناریو ششم	-	۱۱۱۱۱۱۱۱۱۱۱

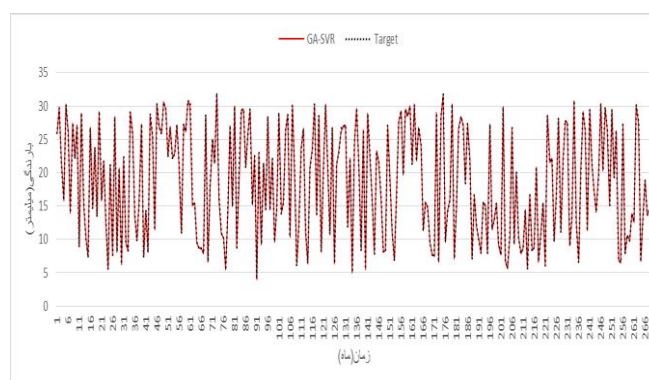
#### نتایج ارزیابی کارایی مدل‌های مورد استفاده

نتایج حاصل از عملکرد دو مدل SVR و GEP به منظور پیش‌بینی بارندگی ماهانه در جدول ۳ و ۴ آورده شده است. در طول دوره آموزش به وسیله SVM مقادیر سه پارامتر  $\epsilon$  و  $\gamma$  با استفاده از سعی و خطا تنظیم گردیده است (جدول ۳). در این سعی و خطا، تلاش بر این بوده که میزان RMSE کمینه و  $R^2$  و NSE نیز به یک نزدیک شوند. مقادیر RMSE،  $R^2$  و NSE مربوط به شش سناریوی موجود نیز در جدول ۳ آورده شده است. در این جدول، C پارامتری است که وظیفه برقراری موازنه میان ساده بودن مدل و میزان خطا را بر عهده دارد  $\epsilon$  میزان اختلاف بین خروجی مطلوب و خروجی مدل، و در واقع میزان خطای قابل تحمل مدل از نظر کاربر می‌باشد. اگر این اختلاف بیشتر از میزان  $\epsilon$  باشد توسط تابع C جریمه خواهد شد. این دو پارامتر بر کیفیت و زمان آموزش تأثیرگذار هستند. مقدار  $\gamma$  نیز بر وقوع کم‌برازش و بیش‌برازش تأثیرگذار است. با توجه به جدول ۳، دقت

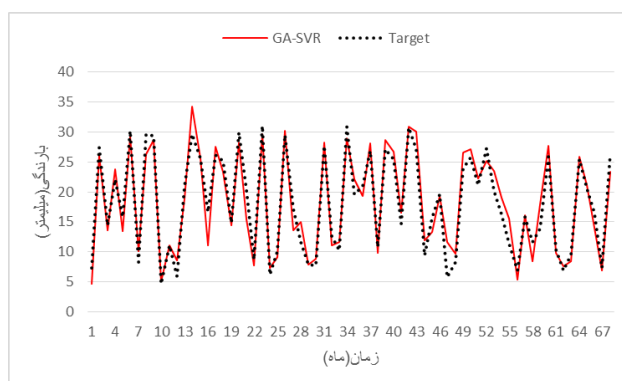
آزمون گاما در انتخاب بهترین ترکیب ورودی محرز است، بطوری که در مرحله اعتبارسنجی سناریوی شاهد که کل ورودی‌ها در نظر گرفته شده است، میزان RMSE بیشتر و مقدار NSE کمتر از سایر سناریوهایی است که ورودی آنها براساس آزمون گاما مشخص شده است. با مقایسه میزان ضرایب جدول ۳ به این نتیجه می‌توان رسید که آزمون گاما در تمام سناریوهایی پیشنهادی خود البته بجز سناریو ۲ از کارایی بهتری نسبت به روش همبستگی (سناریو ۵) برخوردار بوده است. نتایج نشان داد که اگر برای انتخاب ترکیب ورودی از روش همبستگی بین ورودی‌ها و متغیر هدف استفاده شود، عملکرد مدل SVR بهبود می‌یابد. بطوری که در سناریوی ۵ ضرایب  $R^2$  و NSE بیشتر از سناریوی شاهد بوده است. شکل ۵ تا ۸ به ترتیب نمودار پیش‌بینی گاما SVR مربوط به سناریو ۱ و SVR (سناریو شاهد) برای مرحله اعتبارسنجی و آموزش را نشان می‌دهد.

جدول ۳: مقادیر بهینه پارمترها SVR و ارزیابی کارایی آنها برای پیش بینی بارندگی ماهانه ایستگاه شیراز

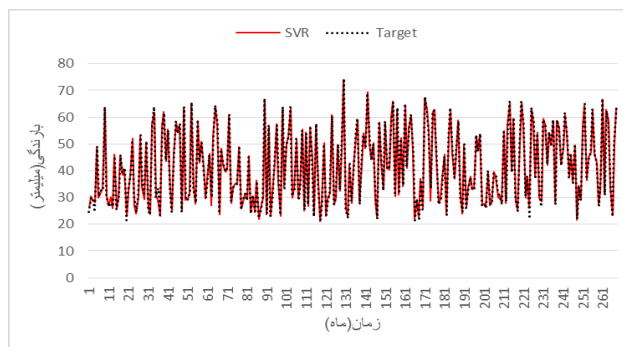
روش	سناریو	آموزش			اعتبارسنجی			SVR (C,γ,ε)
		R <sup>2</sup>	RMSE (mm)	NSE	R <sup>2</sup>	RMSE (mm)	NSE	C,γ,ε
Gamma-SVR	سناریو ۱	۰/۹۹	۰/۰۲	۰/۹۹۸	۰/۹۰	۲/۶۲	۰/۹۲	۰/۰۰۱, ۰/۴, ۱/۴۰
	سناریو ۲	۰/۹۸	۱/۹	۰/۹۸۱	۰/۸۱	۵/۷۶	۰/۸۰۹	۰/۰۰۱, ۰/۴, ۱/۵۰
	سناریو ۳	۰/۹۸	۰/۶۵	۰/۹۹۷	۰/۹۲	۲/۱۴	۰/۸۷۹	۰/۰۰۱, ۰/۴, ۱/۴۰
	سناریو ۴	۰/۹۸	۰/۴۹	۰/۹۹۶	۰/۹۴	۲	۰/۹۴	۰/۰۰۱, ۰/۵, ۱/۴۰
Cor-SVR	سناریو ۵	۰/۹۸	۱/۰۲	۰/۹۹۶	۰/۸۶	۲/۹۷	۰/۶۸	۰/۰۰۱, ۰/۶, ۱/۵۰
SVR	سناریو ۶	۰/۹۸	۱/۱۹	۰/۹۷۴	۰/۷۳	۷/۲۸	۰/۸۹	۰/۰۰۱, ۰/۴, ۱/۴۷



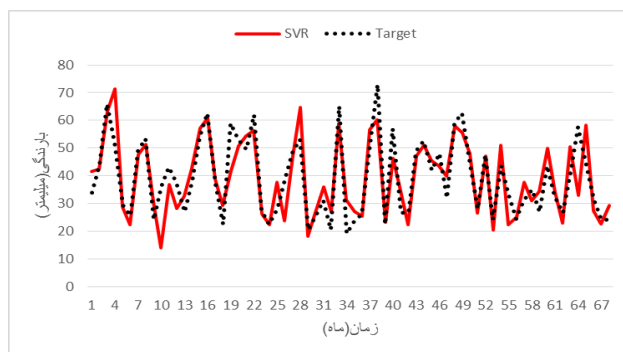
شکل ۵: منحنی پیش بینی شده بارندگی ماهانه با مدل گاما SVR در مرحله آموزش، تحت سناریو ۱



شکل ۶: منحنی پیش بینی بارندگی ماهانه با مدل گاما SVR در مرحله اعتبارسنجی، تحت سناریو ۱



شکل ۷: منحنی پیش‌بینی شده بارندگی ماهانه با مدل SVR در مرحله آموزش، تحت سناریو شاهد



شکل ۸: منحنی پیش‌بینی بارندگی ماهانه با مدل SVR در مرحله اعتبارسنجی، تحت سناریو شاهد

به طور خلاصه پارامترهای استفاده شده در هر بار اجرای مدل GEP در جدول ۴ ارایه شده است.

جدول ۴: مقادیر پارامترهای به کار گرفته شده در برنامه‌ریزی بیان ژن

تعداد کروموزوم‌ها	۳۰	نرخ برگشتی	۰/۱	نرخ ترکیب تک نقطه‌ای	۰/۳
طول هر رأس	۸	نرخ ترانهش ژنی	۰/۱	نرخ ترکیب دو نقطه‌ای	۰/۳
تعداد ژن‌ها	۳	نرخ جهشی	۰/۰۳	نرخ ترکیب ژنی	۰/۱

مورد بررسی در این پژوهش عملکرد بهتری دارد. بنابراین تابع RRSE به عنوان تابع برازش هر سه فاکتور انتخاب شد. گام بعد انتخاب عملگرهای اصلی برای ساخت درخت تجزیه است. توابع ریاضی به کارگرفته شده در این پژوهش  $\{\sqrt{,/,*,+,-, \text{Log}, \text{Exp}, \text{power}, \text{sin}, \text{cos}, \dots}\}$

اولین گام در بررسی مدل برنامه‌ریزی بیان ژن، انتخاب تابع برازش مناسب است که می‌تواند شکل‌های مختلفی به خود بگیرد. نتایج انتخاب اولیه تابع برازش در مدل برنامه‌ریزی بیان ژن بیانگر آن بود که استفاده از تابع برازش ریشه نسبی مربعات خطا ( $RRSE^2$ ) برای همه نمونه‌های

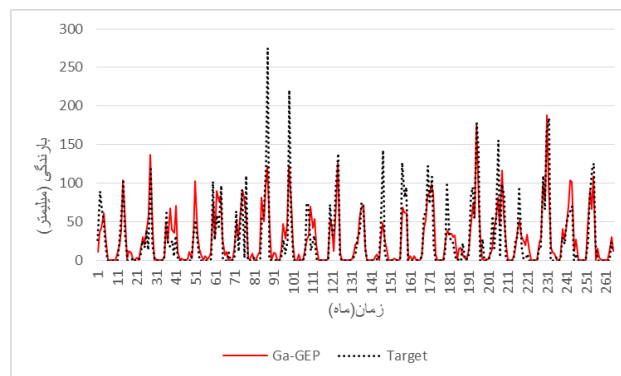
2-Root relative squared

با تابع پیوند ضرب داشت (ابارشی و همکاران، ۱۳۹۳). جدول ۵ عملکرد مدل برنامه‌ریزی بیان ژن در پیش‌بینی بارندگی ماهانه را نشان می‌دهد (شکل ۹ تا ۱۲).

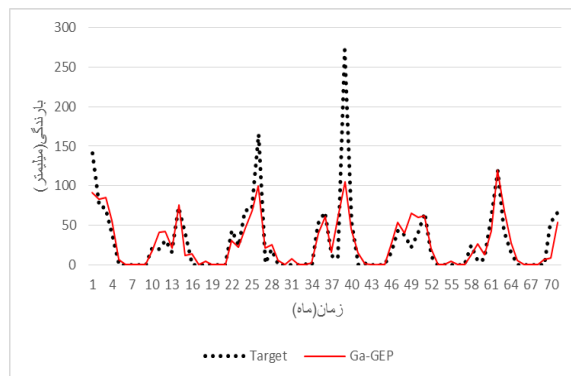
بوده است. پس از انتخاب بهترین تابع برای هر یک از فاکتورها، گام بعد شامل پیدا کردن تابع پیوند مناسب است. از بین توابع پیوند شامل جمع و ضرب، تابع پیوند جمع عملکرد بهتری در مقایسه

جدول ۵: ارزیابی کارایی مدل برنامه‌ریزی بیان ژن برای پیش‌بینی بارندگی ماهانه ایستگاه شیراز

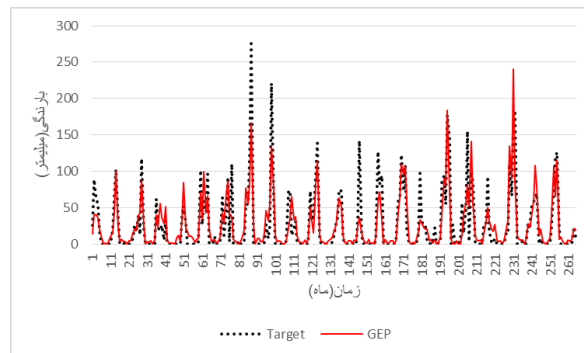
	آموزش	آزمون					
		R <sup>2</sup>	RMSE(mm)	NSE	R <sup>2</sup>	RMSE (mm)	NSE
GA-GEP	سناریوی ۱	۰/۷۱	۲۳/۴	۰/۷۱	۰/۷	۲۵/۷	۰/۶۷
	سناریوی ۲	۰/۷۲	۲۳	۰/۵۵	۰/۶۵	۲۷	۰/۶۳
	سناریوی ۳	۰/۷۲	۲۳/۷	۰/۷	۰/۶۶	۲۷/۹	۰/۶۱
	سناریوی ۴	۰/۷۳	۲۲/۵	۰/۷۳	۰/۶۹	۲۶	۰/۶۶
Cor-GEP	سناریوی ۵	۰/۶۶	۲۶/۲۵	۰/۶۴	۰/۵۶	۳۰/۳	۰/۵۴
GEP	سناریوی شاهد	۰/۷۳	۲۴/۵	۰/۶۸	۰/۸	۲۱/۳	۰/۷۷



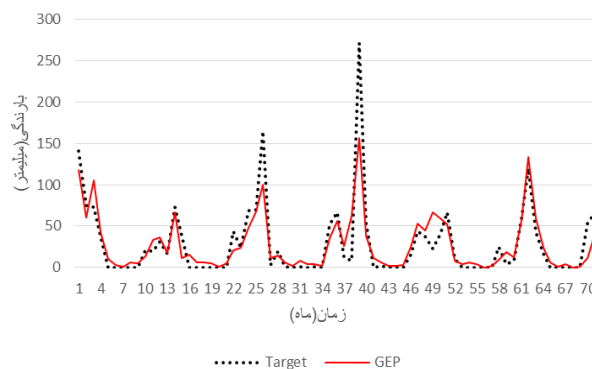
شکل ۹: منحنی پیش‌بینی بارندگی در مرحله آموزش با مدل گاما GEP، تحت سناریو ۱



شکل ۱۰: منحنی پیش‌بینی بارندگی در مرحله آزمون با مدل گاما GEP، تحت سناریو ۱



شکل ۱۱: منحنی پیش‌بینی بارندگی در مرحله آموزش با مدل GEP، تحت سناریو شاهد



شکل ۱۲: منحنی پیش‌بینی بارندگی در مرحله آزمون با مدل GEP، تحت سناریو شاهد

مقایسه روش‌های گاما و همبستگی برای انتخاب ترکیب ورودی مناسب (جدول ۵) مشخص می‌شود که براساس میزان ضرایب NSE، RMSE و  $R^2$  روش گاما عملکرد بهتری را داشته است. از طرف دیگر، با مقایسه سناریوی ۵ و شاهد، انتخاب ترکیب ورودی به روش همبستگی نیز مانند روش گاما قادر به بهبود کارایی مدل GEP نبوده است. شکل ۹ تا ۱۲ نمودارهای مربوط به پیش‌بینی مدل GEP را براساس سناریوهای شماره ۱ و شاهد را در مرحله آموزش و اعتبارسنجی نشان می‌دهد.

#### بحث

شاخص‌های آماری مربوط به نتایج حاصل از کاربرد روش‌های رگرسیون بردار پشتیبان و برنامه‌ریزی

نتایج جدول ۵ نشان می‌دهد که در سناریو شماره ۱ که شامل متغیرهای (H, E, Sh, Sh-1, Wv-1) می‌باشد، میزان RMSE حاصل از مدل GEP در مرحله آزمایش برابر ۲۵/۷ میلیمتر است، که از مقادیر RMSE حاصل شده از ۳ سناریو بعدی گاما و سناریوی ۵ (ترکیب متغیرهای که بیشترین همبستگی را با متغیر هدف دارا بوده‌اند شامل T, H, E و Sh) کمتر، اما نسبت به سناریو شاهد دارای عملکرد ضعیف‌تری بوده است. نتایج ضریب کارایی NSE نیز مشابه با RMSE بوده است. در سناریوی شاهد، میزان ضریب  $R^2$  مدل GEP در هردو حالت آموزش و اعتبارسنجی، بیشتر از سناریوهای یک تا چهار که ورودی‌های آنها براساس آماره گاما مشخص شده، می‌باشد. با

شیری (۲۰۱۱) و به نقل از بناف و همکاران (۱۹۹۸) دلیل این موضوع را اینگونه بیان می‌کنند که متغیرهای ورودی انتخاب شده، بصورت کامل محیط را برای یادگیری سیستم فراهم نخواهند کرد، و دامنه یادگیری براساس انتخاب‌های انجام شده گذشته باید باشد. آنها معتقد هستند که انتخاب هر متغیر ورودی حاوی یکسری اطلاعات در ارتباط با متغیر خروجی است که در آموزش مدل از آنها استفاده می‌شود و در صورت تناسب داشتن توسط خود مدل GEP برای یادگیری سیستم انتخاب می‌گردد. نکته قابل توجه این است که در مدل GEP در هر دو حالت شاهد و سناریوی ۱، تخمین تعداد زیادی از مقادیر حد بالا همواره با خطا همراه بوده است. در حالی که روش SVR میزان خطا کمتری در تخمین مقادیر اوج را دارا بوده است، که دلیل آنرا می‌توان گرفتار نشدن آن در حداقل‌های محلی دانست (مینکانگ و همکاران، ۲۰۱۰). با توجه به اینکه خطا در تخمین مقادیر حدی برای تمام مدل‌ها GEP و ساختارهای متفاوت آن رخ داده است پس نمی‌توان به دلیل ساختار آنها باشد و باید دنبال دلایل دیگری برای آن بود. دلیل محتمل برای این نقص را می‌توان به استفاده از اطلاعات کم بارش حداقل و حداکثر قبل از زمان پیش‌بینی مربوط کرد، که همانند شبکه‌های عصبی مصنوعی مدل GEP نیز می‌تواند در صورت آشنا نشدن با روند تغییرات بارش در مرحله آموزش دچار اختلال در تخمین مقادیر حدی شود (خلیلی و همکاران، ۱۳۸۶). بررسی فراوانی مقادیر مشاهداتی بارندگی نشان داد که ۵۸٪ داده‌ها در بازه بین ۰ تا ۱۰ میلیمتر قرار دارند، پس دلیل محتمل تعداد کم داده‌های حد بالا در قسمت آموزش می‌تواند قابل اتکا باشد. ذکر این نکته الزامی است که انتخاب داده‌های آموزش بصورت تصادفی صورت گرفته است. آزمون گاما

بیان ژن در شبیه‌سازی تغییرات بارندگی در جداول ۳ و ۵ ارائه گردیده است. از مقایسه این جداول چنین بر می‌آید که در تمام سناریوهای ورودی چه در مرحله آموزش و چه در مرحله آزمون، رگرسیون بردار پشتیبان نسبت به برنامه‌ریزی بیان ژن از صحت بیشتر و خطای کمتری برخوردار است. مدل SVR برای سناریوی شاهد که تمام متغیرهای ورودی در آن وجود دارند نسبت به GEP عملکرد بهتری را نشان داده است. تفاوت عملکردی مدل SVR در هنگام انتخاب ترکیب‌های مختلف از متغیرهای ورودی چندان نیست، که این نشان دهنده تأثیرپذیری کم این روش از تعداد ورودی‌ها است. در صورتی که در مدل GEP تفاوت عملکردی بین بهترین سناریو (شماره ۱) و سناریو شاهد محسوس می‌باشد. از طرفی نتایج نشان داد که کارایی هر دو مدل انتخابی زمانی که انتخاب متغیرهای ورودی با استفاده از آزمون گاما صورت می‌گیرد به مراتب بهتر از حالتی است که انتخاب متغیرهای ورودی با روش همبستگی صورت گرفته باشد (احمدی و همکاران، ۲۰۱۴). با مقایسه عملکرد مدل SVR در حالتی که کل متغیرها به‌عنوان ترکیب ورودی (سناریو شاهد) آن در نظر گرفته شده است و حالتی که پیش‌پردازش بر روی داده‌ها صورت گیرد، مشخص شد که کارایی این مدل بسته به نوع پیش‌پردازش انجام شده می‌تواند بهبود یابد. اما در روش برنامه‌ریزی بیان ژن، خواه با آزمون گاما و یا همبستگی، بهبودی در نتایج مدل ایجاد نشده است. به عبارتی دیگر، در زمان استفاده از سناریوی شاهد (وجود کلیه متغیرها) روش برنامه‌ریزی بیان ژن بهترین نتیجه را رقم زده است. پس این امر نشان می‌دهد که برخلاف نتایج قربانی و همکاران (۱۳۹۲) افزایش حافظه در بهبود دقت مدل GEP موثر بوده است. کیسی و



### نتیجه‌گیری

در تخمین بارندگی به کمک روش‌های یادگیری ماشینی انتخاب ترکیب متغیرهای ورودی مورد سوال بوده است که در این تحقیق آزمون گاما برای این منظور مورد استفاده قرار گرفت. براساس نتایج حاصل از بکارگیری آزمون گاما، متغیرهای رطوبت نسبی (H)، تبخیر (E)، تعداد ساعات آفتابی (Sh)، تعداد ساعات آفتابی یک ماه قبل (Sh-1)، سرعت باد یک ماه قبل (Wv-1) به‌عنوان بهترین ترکیب ورودی مدل‌های مورد استفاده در این تحقیق انتخاب شدند. در مجموع، مطالعه انجام شده آزمون گاما را به‌عنوان روشی مناسب برای انتخاب متغیرهای تأثیر گذار (محلی) بر پیش‌بینی بارندگی ماهانه با استفاده از مدل ماشین بردار پشتیبان معرفی می‌کند. براساس نتایج حاصله از بکارگیری مدل رگرسیون بردار پشتیبان، براساس بهترین ترکیب ورودی مدل مبتنی بر آزمون گاما، مقادیر معیارهای آماری ارزیابی کارایی  $R^2$ ، RMSE و NSE برای آموزش مدل به ترتیب ۰/۹۹، ۰/۰۲ و ۰/۹۹۸ و برای اعتبارسنجی مدل به ترتیب ۰/۹، ۲/۶۲ و ۰/۹۲ بدست آمد. هم‌چنین برای مدل رگرسیون بردار پشتیبان، براساس بهترین ترکیب ورودی مدل مبتنی بر آزمون همبستگی، مقادیر معیارهای آماری ارزیابی کارایی  $R^2$ ، RMSE و NSE برای آموزش مدل به ترتیب ۰/۹۸، ۱/۰۲ و ۰/۹۹۶ و برای اعتبارسنجی مدل به ترتیب ۰/۸۶، ۲/۹۷ و ۰/۶۸ برآورد گردید که مقایسه نتایج ارزیابی کارایی مدل رگرسیون بردار پشتیبان با انتخاب بهترین ترکیب ورودی مبتنی بر آزمون گاما و آزمون همبستگی نشان می‌دهد که بکارگیری آزمون گاما موجب کارایی بهتر مدل رگرسیون بردار پشتیبان نسبت به بکارگیری آزمون همبستگی می‌گردد. علاوه بر این، براساس نتایج حاصله از بکارگیری مدل برنامه‌ریزی بیان

متغیرهای ساعات آفتابی و رطوبت نسبی را به‌عنوان اصلی‌ترین متغیرها در پیش‌بینی بارندگی شیراز معرفی کرده است که با نتایج لین و همکاران (۲۰۱۳) مطابقت دارد. به طوری که در سناریوی ۱ که SVR بهترین عملکرد را داشته و چهار سناریوی دیگر نیز این دو متغیر همواره وجود داشته‌اند.

در این تحقیق برای تعیین طول دوره آموزش از روش آزمون M استفاده شد. ارزیابی کارایی این روش، با استفاده از تغییر تعداد داده‌های آموزش و بررسی خطای آنها انجام شد. در بازه  $5 \pm / 85$  داده‌ها تغییرات ملموسی در نتایج هیچکدام از مدل‌ها ایجاد نگردیده، ولی با عبور از این بازه کارایی مدل‌ها کاهش یافته است. به این دلیل در این تحقیق ۸۰ درصد داده‌ها به‌عنوان طول دوره آموزش انتخاب شده است (نوری، ۲۰۰۹). این تحقیق دقت پیش‌بینی بارندگی را با استفاده از متغیرهای اقلیمی محلی شامل ساعات آفتابی، رطوبت نسبی، دما، تبخیر و سرعت باد از طریق مدل گاما SVR به مراتب بهتر از نتایج تحقیقات مشابه انجام داده است (ویو و همکاران، ۲۰۱۰)، اما نتایج این تحقیق در بهبود کارایی روش برنامه‌ریزی بیان ژن در پیش‌بینی بارندگی نسبت به تحقیقات مشابه انجام شده قبلی ناتوان بوده است. به‌عنوان مثال ستار و نهرین (۱۳۹۲) حداکثر بارش روزانه ایستگاه هواشناسی اهر و جلفا را با استفاده از سه روش شبکه عصبی مصنوعی، برنامه‌ریزی ژنتیک و مدل درختی M5 پیش‌بینی کردند که نتایج تحقیق آنها به مراتب دارای RMSE کمتری نسبت به روش برنامه‌ریزی بیان ژن در این تحقیق بوده است. کیسی و شیرینی (۲۰۱۱) با ترکیب موجک با برنامه‌ریزی بیان ژن دقت پیش‌بینی بارندگی در دو شهر ازمیر و آفیون ترکیه را بهتر از تحقیق حاضر انجام داده‌اند.

پیش‌بینی بارندگی ماهانه ایستگاه شیراز با استفاده از دو مدل رگرسیون بردار پشتیبان و برنامه‌ریزی بیان ژن، کارایی برتر رگرسیون بردار پشتیبان نسبت به برنامه‌ریزی بیان ژن را نشان می‌دهد. بطور کلی مدل رگرسیون بردار پشتیبان را می‌توان برای محاسبات دقیق‌تر در زمینه حل مسائل مربوط به منابع آب بخصوص پیش‌بینی بارندگی استفاده کرد. به منظور کسب نتایج بهتر، برای تحقیقات آینده پیشنهاد می‌شود که از سیگنال‌های بزرگ مقیاس اقلیمی به‌عنوان ورودی مدل GEP استفاده شود تا کارایی این روش در حالت‌های متفاوت‌تری هم مورد بررسی قرار گرفته باشد. همچنین پیشنهاد می‌گردد که پیش‌پردازش‌های دیگر از جمله تحلیل موجک برای بهبود نتایج استفاده گردد. در این تحقیق فقط از یک ایستگاه سینوپتیک استفاده شده و پیشنهاد می‌گردد که از ایستگاه‌های بیشتری استفاده شود.

کشاورزی (ویژه آب و خاک)، جلد ۲۲، شماره ۱، ص ۸۹-۹۹.

-دستورانی، م.ت.، عظیمی فشی، خ.، طالبی، ع. و اختصاصی، م.ر.، ۱۳۹۱. برآورد رسوبات معلق با استفاده از شبکه عصبی مصنوعی (مطالعه موردی: حوزه آبخیز جامیشان استان کرمانشاه)، پژوهشنامه مدیریت حوزه آبخیز دانشگاه علوم کشاورزی و منابع طبیعی ساری، سال سوم، پائیز و زمستان، شماره ۶، ص ۶۱-۷۴.

-سلطانی، ع.، قربانی، م. ع.، فاخری فرد، ا.، دربندی، ص. و فرسادی زاده، د.، ۱۳۸۹. برنامه‌ریزی ژنتیک و کاربرد آن در مدل‌سازی فرآیند بارش-رواناب، مجله دانش آب و خاک، جلد ۲۰، شماره ۴، ص ۶۱-۷۱.

-شریفی، ع.، دین‌پژوه، ی.، فاخری فرد، ا. و مقدم نیا، ع.، ۱۳۹۳. ترکیب بهینه متغیرها برای

ژن، براساس بهترین ترکیب ورودی مدل مبتنی بر آزمون گاما، مقادیر معیارهای آماری ارزیابی کارایی  $R^2$ ، RMSE و NSE برای آموزش مدل به ترتیب ۰/۷۱، ۲۳/۴ و ۰/۷۱ و برای اعتبارسنجی مدل به ترتیب ۰/۷، ۲۵/۷ و ۰/۶۷ بدست آمد. همچنین برای مدل برنامه‌ریزی بیان ژن، براساس بهترین ورودی مدل مبتنی بر آزمون همبستگی، مقادیر معیارهای آماری ارزیابی کارایی  $R^2$ ، RMSE و NSE برای آموزش مدل به ترتیب ۰/۶۶، ۲۶/۲۲۵ و ۰/۶۴ و برای اعتبارسنجی مدل به ترتیب ۰/۵۶، ۳۰/۳ و ۰/۵۴ برآورد گردید که مقایسه نتایج ارزیابی کارایی مدل برنامه‌ریزی بیان ژن با انتخاب بهترین ترکیب ورودی مبتنی بر آزمون گاما و آزمون همبستگی نشان می‌دهد که بکارگیری آزمون گاما موجب کارایی بهتر مدل برنامه‌ریزی بیان ژن نسبت به بکارگیری آزمون همبستگی می‌شود. در نهایت، مقایسه نتایج بدست آمده از

## منابع

-بارش‌ی، ف.، مفتاح هلقی، م.، ثانی‌خانی، ه. و دهقانی، ا.، ۱۳۹۳. مقایسه عملکرد سه روش هوشمند به منظور پیش‌بینی نوسانات سطح ایستابی، مطالعه موردی: دشت زرین گل، نشریه پژوهش‌های حفاظت آب و خاک، جلد ۲۱، شماره ۱، ص ۱۶۳-۱۸۰.

-احمدی، آ.، مریدی، ع.، کاکایی لعدانی، ا. و کیان پیشه، ق.، ۱۳۹۳. پیش‌بینی بلند مدت بارش بر پایه الگوهای پیوند دور اقلیمی، مطالعه موردی: حوضه اهرچای، فصلنامه علمی-پژوهشی آب و فاضلاب، جلد ۲۵، شماره ۵، ص ۸۶-۹۶.

-خلیلی، ن.، خدائیاناس، س.ر. و داوری، ک.، ۱۳۸۷. پیش‌بینی بارش ماهانه با استفاده از شبکه‌های عصبی مصنوعی (مطالعه موردی ایستگاه سینوپتیک مشهد)، مجله علوم و صنایع

در مدل‌سازی بارندگی فصلی (مطالعه موردی: ایستگاه‌های منتخب استان خوزستان)، فصلنامه تحقیقات جغرافیایی، جلد ۲۹، شماره ۱۱۴، ص ۶۱-۷۲.

شبهه‌سازی رواناب در حوزه آبخیز امامه با استفاده از آزمون گاما، دانش آب و خاک، جلد ۲۳، شماره ۴، ص ۵۹-۷۲.

-گلابی، م.، آخوندعلی، ع.م. و رادمنش، ف.، ۱۳۹۳. مقایسه دقت پیش‌بینی مدل‌های باکس-جنکینز

-Ahmadi, A., Han, D., Karamouz, M. and Remesan, R., 2009. Input data selection for solar radiation estimation, *Hydrological processes*, v. 23(19), p. 2754-2764, doi: 10.1002/hyp.7372.

-Ahmadi, A., Moridi, A., Lafdani, E.K. and Kianpisheh, G., 2014. Assessment of climate change impacts on rainfall using large scale climate variables and downscaling models-A case study, *Journal of earth system science*, v. 123(7), 1603 p.

-Ahmadi, A., Han, D., Lafdani, E. K. and Moridi, A., 2015. Input selection for long-lead precipitation prediction using large-scale climate variables: a case study, *Journal of Hydroinformatics*, v. 17(1), p. 114-129.

-Chuan, C.S., 1997. weather prediction using artificial neural network, *Journal of Hydrology*, v. 230, p. 101-119.

-Chang, F. J., Tsai, Y. H., Chen, P. A., Coynel, A. and Vachaud, G., 2015. Modeling water quality in an urban river using hydrological factors-Data driven approaches, *Journal of environmental management*, v. 151, p. 87-96.

-Cristianini, N. and Shawe-Taylor, J., 2000. An introduction to support vector machines and other kernel-based learning methods, Cambridge university press, 189 p.

-El-Shafie, A., Mukhlisin, M., Najah, A.A. and Taha, M.R., 2011. Performance of artificial neural network and regression techniques for rainfall-runoff prediction, *International Journal of Physical Sciences*, v. 6(8), p. 1997-2003.

-Ferreira, C., 2006. Gene expression programming: mathematical modeling by an artificial intelligence, 2nd ed, Springer-Verlag, Germany, 478 p.

-Ghabaei, S.M., Mosaedi, A., Hesam, M. and Hezarjaribi, A., 2010. Evaluation effect of input parameters preprocessing in artificial neural networks (Anns) by using stepwise regression and Gamma Test techniques for fast estimation of daily evapotranspiration, p 1-14.

-Karamouz, M., Fallahi, M., Nazif, S. and Rahimi Farahani, M., 2009. Long lead rainfall prediction using statistical downscaling and artificial neural network modeling, *Scintia Iranica*, v. 16(1), p. 165-72.

-Kisi, O. and Cimen, M., 2011. A wavelet-support vector machine conjunction model for monthly streamflow forecasting, *Journal of Hydrology*, v. 399(1), p. 132-140.

-Lin, G.F., Jhong, B.C. and Chang, C.C., 2013. Development of an effective data-driven model for hourly typhoon rainfall forecasting, *Journal of Hydrology*, v. 495, p. 52-63.

-Lopes, H.S. and Weinert, W.R., 2004. EGIPSY: an enhanced gene expression programming approach for symbolic regression problems, *International Journal of Applied Mathematics and Computer Science*, v. 14(3), p. 375-384.

-Moghaddamnia, A., Ghafari, M., Piri, J. and Han, D., 2009. Evaporation estimation using support vector machines technique, *International*

Journal of Engineering and Applied Sciences, v. 5(7), p. 415-423.

-Najafi, M.R., Moradkhani, H. and Wherry, S.A., 2010. Statistical downscaling of precipitation using machine learning with optimal predictor selection, Journal of Hydrologic Engineering, v. 16(8), p. 650-664.

-Noori, R., Abdoli, M.A., Ghasrodashti, A.A. and Jalili Ghazizade, M., 2009. Prediction of municipal solid waste generation with combination of support vector machine and principal component analysis: a case study of Mashhad, Environmental Progress & Sustainable Energy, v. 28(2), p. 249-258.

-Patil, C.Y. and Ghatol, A.A., 2010. Rainfall forecasting using local parameters over a meteorological station: an artificial neural network approach, International Journal of Engineering Research & Industrial Applications, v. 3, p. 341-356.

-Rabunal, J.R., Puertas, J., Suarez, J. and Rivero, D., 2007. Determination of the unit hydrograph of a typical urban basin using genetic programming and artificial neural networks, Hydrological processes, v. 21(4), p. 476-485.

-Vapnik, V., 2013. The nature of statistical learning theory, Springer Science & Business, 314 p.

-Weerasinghe, H.D.P., Premaratne, H.L. and Sonnadara, D.U.J., 2010. Performance of neural networks in forecasting daily precipitation using multiple sources, Journal of the National Science Foundation of Sri Lanka, v. 38(3), p. 163-170.

-Wu, C. L., Chau, K.W. and Fan, C., 2010. Prediction of rainfall time series using modular artificial neural networks coupled with data-preprocessing techniques, Journal of Hydrology, v. 389(1), p. 146-167.