

ارزیابی عملکرد روش‌های انتخاب متغیر در ریزمقیاس نمایش بارش روزانه دو اقلیم متفاوت

جواد رضوانی مقدم^{۱*} - مصطفی یعقوب زاده^۲ - احمد جعفرزاده^۳

تاریخ دریافت: ۱۳۹۷/۰۳/۰۱

تاریخ پذیرش: ۱۳۹۷/۰۵/۰۷

چکیده

فرآیند ریزمقیاس نمایش آماری با هدف ارتقای شبیه‌سازی‌های مدل‌های GCMs و کاربست نتایج آن‌ها در مقیاس محلی انجام می‌شود. در این بین انتخاب متغیرهای ورودی ریزمقیاس نمایش اولین گام مهم این فرآیند می‌باشد. از آنجا که هدف اصلی ریزمقیاس نمایش بهبود شبیه‌سازی مدل‌های اقلیمی می‌باشد، بسیاری از مطالعات روش‌های متنوعی را برای انتخاب متغیرهای ورودی ریزمقیاس نمایش مورد ارزیابی قرار داده‌اند. این مطالعه در نظر دارد تا با استفاده از آزمون‌های مقایسه‌ای جامع، عملکرد شبکه عصبی را در فرآیند ریزمقیاس نمایش بارش روزانه تحت تأثیر چهار روش انتخاب متغیر PCA، SRA، CA و ParCA در اقلیم‌های متفاوت مورد ارزیابی قرار دهد. بدین منظور در ابتدا داده‌های مشاهداتی ۳۰ ساله مربوط به ایستگاه‌های بیرجند (اقلیم خشک- کویری) و اردبیل (اقلیم سرد- نیمه‌خشک)، حدفاصل سال‌های ۲۰۰۴-۱۹۷۷ گردآوری شد. به منظور شبیه‌سازی رفتار مؤلفه‌های اقلیمی متأثر از پدیده تغییر اقلیم از خروجی مدل CanESM2 استفاده شد. بدین ترتیب داده‌های بزرگ مقیاس مدل CanESM2 برای هر دو ایستگاه سینوپتیک به منزله متغیرهای ورودی و بارندگی مشاهداتی به عنوان متغیر خروجی در نظر گرفته شد. آزمون‌های مقایسه‌ای شامل شاخص‌های ارزیابی، مقایسه مشخصه‌های آماری، جدول Contingency Table Event جهت تشخیص سری روزهای تر و خشک و مقایسه نموداری توزیع آماری از جمله ابزارهای مورد استفاده در این مطالعه جهت ارزیابی عملکرد روش‌های مختلف انتخاب متغیر می‌باشد. نتایج مطالعه نشان داد که به طور کلی ریزمقیاس نمایش بارش روزانه در تمامی روش‌های انتخاب متغیر در ایستگاه بیرجند دارای عملکرد بهتری نسبت به ایستگاه اردبیل می‌باشد. همچنین نتایج آزمون‌های مختلف نشان داد که روش‌های انتخاب متغیر CA و ParCA در اقلیم‌های خشک و روش SRA در اقلیم سرد-نیمه‌خشک از عملکرد بهتری برخوردار می‌باشد. بهترین مقادیر شاخص‌های RMSE، R و NSE برای ایستگاه بیرجند به ترتیب ۱/۲ میلی‌متر در روز، ۰/۵۵ و ۰/۲۵ و در ایستگاه اردبیل به ترتیب ۱/۷۵ میلی‌متر در روز، ۰/۱۴ و ۰/۱۳ بدست آمد. ارزیابی روش‌ها در تشخیص درست روزهای تر و خشک در بیرجند نشان داد که دقت روش‌های CA و ParCA به ترتیب ۲۵ و ۲۲ درصد می‌باشد. این بدان معنی است که روش CA توانسته است ۲۵ روزهای تر را به درستی تر تشخیص دهد.

واژه‌های کلیدی: تحلیل مؤلفه‌های اصلی، دوره‌های تر و خشک، کاهش ابعاد ورودی، همبستگی جزئی و CanESM2

مقدمه

رفتار مؤلفه‌های اقلیمی متأثر از پدیده تغییر اقلیم از آن یاد می‌شود، خروجی مدل‌های GCMs^۴ می‌باشد. این مدل‌ها عمدتاً در مراکز هواشناسی و مطالعات فضایی کشورهای توسعه یافته تدوین شده‌اند. مرکز IPCC هر ۴ یا ۵ سال با استفاده از خروجی این مدل‌ها تغییرات مؤلفه‌های اقلیمی را در سطح کره زمین و در قالب گزارش‌های ارزیابی ارائه می‌کند. این مدل‌ها سطح کره زمین را به صورت نقاط شبکه‌بندی شده مدل‌سازیمی‌کنند و خروجی‌های خود را در مرکز این نقاط ارائه می‌کنند. فاصله این نقاط شبکه‌بندی شده که نشان دهنده دقت مدل‌های GCMs می‌باشد بسیار زیاد بوده و برای شبیه‌سازی رفتار مؤلفه‌های اقلیم در مقیاس محلی و ایستگاهی مناسب نمی‌باشند. به منظور بررسی اثرات تغییر اقلیم بر منابع آب در یک منطقه خاص

در حال حاضر شواهد زیادی مبنی بر وقوع پدیده تغییر اقلیم و پیامدهای آن وجود دارد. IPCC^۳ در گزارش‌های ارزیابی خود افزایش دمای کره زمین را بیان نموده است (۱۶، ۱۷ و ۳۴). از طرفی دیگر مهم‌ترین و مطمئن‌ترین ابزاری که در مجامع علمی برای شبیه‌سازی

۱- استادیار گروه مهندسی آب، دانشکده کشاورزی و منابع طبیعی، دانشگاه محقق اردبیلی، اردبیل، ایران
(*) نویسنده مسئول: (Email: j_ramezani@uma.ac.ir)
۲ و ۳- استادیار و دانشجوی دکتری مهندسی منابع آب، گروه علوم و مهندسی آب، دانشکده کشاورزی، دانشگاه بیرجند

بسیاری از پژوهش‌های انجام شده برای انتخاب متغیر از آزمون CA^5 استفاده نموده‌اند (۱۳ و ۲۷). به عنوان مثال مطالعه چن و همکاران (۴) که با هدف بررسی پیامدهای تغییر اقلیم بر روی رواناب حوضه بالادست هانجیانگ^۶ کشور چین انجام شده است، جهت انتخاب متغیرهای ورودی^۷ SVM^۸ و SDSM^۹ از روش CA استفاده نموده‌اند (۴). همچنین در مطالعه مینیو و همکاران (۲۴) نیز برای انتخاب متغیرهای ورودی به SDSM جهت ریزمقیاس‌نمایی بارش روزانه و ارزیابی پیامدهای تغییر اقلیم بر حوضه آبریز تونگاپهادرا (Tungabhadra) کشور هندوستان از روش CA بهره گرفته شده است (۲۴). در مطالعه ناصری و همکاران (۲۶) که با هدف ارزیابی روش‌های ریزمقیاس‌نمایی در شبیه‌سازی بارش روزانه انجام شد، متغیرهای ورودی جهت ریزمقیاس‌نمایی با استفاده از روش ParCA^{۱۰} انتخاب شدند. در پژوهش آن‌ها که از خروجی مدل HadCM3 تحت دو سناریوی A2 و B2 به منظور شبیه‌سازی بارش روزانه بهره گرفته شده بود، متغیرهای رطوبت نسبی و سرعت ناحیه‌ای در ارتفاع‌های متفاوت ژئوپتانسیل به عنوان متغیر ورودی برگزیده شدند (۲۶). روش رگرسیون گام به گام^{۱۱} SRA بعد از روش CA بیشترین کاربرد را در مطالعات مربوط به ریزمقیاس‌نمایی خروجی مدل‌های GCMs دارا می‌باشد. روش‌های مبتنی بر رگرسیون عموماً به دنبال پیدا نمودن مجموعه‌ای از متغیرهای مستقل می‌باشند که مجذور خطای معادله رگرسیونی آن‌ها با متغیر وابسته حداقل شود. خاشعی و همکاران (۲۱) و پژوهش‌های هات (۱۵)، حسامی و همکاران (۱۴) و چن و همکاران (۵) از جمله تحقیقاتی می‌باشند که جهت انتخاب متغیرهای ورودی به ریزمقیاس‌نمایی از روش SRA استفاده نمودند (۲۱ و ۵). مطالعه حسامی و همکاران (۱۴) به منظور شبیه‌سازی مؤلفه‌های حدی در شرق کانادا با استفاده از خروجی مدل گردش عمومی CGCMS1، از ریزمقیاس‌نمایی SDSM بهره گرفتند. آن‌ها در مطالعه خود جهت انتخاب متغیرهای ورودی به SDSM از روش SRA استفاده نمودند (۱۴). همچنین مطالعه چن و همکاران (۱۴) به منظور بررسی اثرات تغییر اقلیم بر دبی رودخانه‌های هی^{۱۱} چین، از خروجی مدل HadCM3 تحت سناریوهای A2 و B2 استفاده نمودند. روش ریزمقیاس‌نمایی آن‌ها SDSM و متوسط درجه حرارت، تأخیر و بارندگی به عنوان متغیرهای اقلیمی مورد ارزیابی واقع شدند (۵). همچنین بعضی از مطالعات نظیر هارفام و

لازم است تا خروجی این مدل‌ها در مقیاس محلی و یا منطقه‌ای تبدیل گردد (۳۱). به تبدیل شبیه‌سازی مدل‌های GCMs از مقیاس‌های جهانی به محلی، ریزمقیاس‌نمایی گفته می‌شود. در حالت کلی ریزمقیاس‌نمایی در دو دسته کلی تقسیم‌بندی می‌شود: آماری و دینامیکی. روش دینامیکی بر پایه شبیه‌سازی مدل‌های RCM^۱ می‌باشد. مدل‌های RCM در مقیاسه با مدل‌های GCMs دارای دقت بالاتری می‌باشند. این بدین جهت می‌باشد که این مدل‌ها خروجی مدل‌های GCMs را بهبود داده و شبیه‌سازی‌های خود را برای یک منطقه خاص پیاده‌سازی کرده‌اند. البته این دقت بالای ریزمقیاس‌نمایی دینامیکی در کنار هزینه و زمان بیشتر، نیاز به فناوری و صنعت بالا دارد (۲ و ۳۰). این محدودیت‌ها باعث شده است که عمده پژوهش‌های مربوط به بررسی پیامدهای تغییر اقلیم جهت ریزمقیاس‌نمایی متوجه روش‌های آماری باشد. روش‌های آماری با ایجاد یک رابطه بین خروجی مدل‌های GCMs و متغیرهای اقلیمی عمل می‌کنند (۳۶). روش‌های آماری به سه دسته مولدهای اقلیمی^۲، توابع انتقال^۳ و الگوهای هواشناسی^۴ تقسیم می‌شوند. به دلیل توسعه برنامه‌های رایانه‌ای و افزایش قدرت شبیه‌سازی مبتنی بر برنامه‌نویسی، توابع انتقال توجه بخش زیادی از پژوهش‌ها را به خود اختصاص داده است. توابع انتقال با استفاده از رگرسیون‌های خطی و غیرخطی، در پی مدل‌سازی بین متغیرهای بزرگ مقیاس (خروجی مدل‌های GCMs) و متغیرهای کوچک مقیاس محلی می‌باشند (۶). در ریزمقیاس‌نمایی به روش توابع انتقال خروجی مدل‌های GCMs به عنوان ورودی و مؤلفه‌های اقلیمی به عنوان خروجی در نظر گرفته می‌شود. در این بین آنچه که حایز اهمیت است، انتخاب متغیرهای بزرگ مقیاس مناسب جهت پیدا نمودن رابطه آماری با متغیرهای محلی می‌باشد. در تحقیقات مربوط به ارزیابی پیامدهای تغییر اقلیم در حوضه منابع آب که مبتنی بر روش‌های توابع انتقال می‌باشد، اولین گامانتخاب متغیرهای ورودی یا پیش‌بینی‌کننده‌ها می‌باشد. انتخاب متغیرهای ورودی در مطالعات گذشته بر اساس دانش و آگاهی محقق انجام می‌شد و اخیراً بر اساس روابط آماری موجود بین متغیرهای ورودی و خروجی می‌باشد. در ادامه سعی شده است تا ضمن مروری جامع، اکثر روش‌های انتخاب متغیر در مطالعات ریزمقیاس‌نمایی متغیرهای اقلیمی مورد ارزیابی قرار گیرد. بررسی مطالعات انجام شده در مقوله ریزمقیاس‌نمایی نشان می‌دهد که به دلیل این‌که یکی از مهم‌ترین معیارهای انتخاب متغیرهای ورودی برای فرآیند ریزمقیاس‌نمایی، داشتن همبستگی بسیار قوی با داده‌های اقلیمی می‌باشد (۳۳)،

5-Correlation Analysis

6-Hanjiang

7-Support Vector Machine

8-Statistical Downscaling Model

9-Partial Correlation Analysis

10-Stepwise Regression Analysis

11-Haihe

1- Regional climate model

2-Weather Generator

3-Transfer Function/ Data Driver/ Black Box

4-Weather Pattern

و ویلای (۱۲) با استفاده از چندین روش ترکیبی سعی نموده‌اند تا بهترین متغیرهای ورودی را جهت ریزمقیاس نمایی انتخاب نمایند. آن‌ها در مطالعه خود از روش‌های SRA^۱، GA^۱ و یک روش مبتنی بر فاصله اطمینان، متغیرهای ورودی شبکه عصبی (Artificial Neural Network-ANN) و SDSM را جهت ریزمقیاس نمایی بارش روزانه انتخاب نمودند (۱۲). همچنین فیستوگلو و اوکان (۹) در مطالعه خود براساس شاخص‌های NSE^۲ و ضریب ملو^۳ تمام ترکیب‌های ممکن برای ورود به ریزمقیاس نمایی بارش روزانه را رتبه‌دهی نمودند و در نهایت ترکیبی که بیشترین امتیاز را داشت به عنوان ورودی‌های ریزمقیاس نمایی انتخاب نمودند (۹). دسته‌ای دیگر از مطالعات ریزمقیاس نمایی خروجی مدل‌های گردش عمومی نیز می‌باشند که از روش‌های متفاوتی نظیر PCA^۴، Gamma Test و ICA^۵ جهت انتخاب متغیرهای ورودی استفاده نموده‌اند. به عنوان مثال کاویانی و میررکنی (۱۹) در پژوهش خود به منظور پیش‌بینی دمای میانگین روزانه ایستگاه همدیدی شهرستان یزد از روش ANN استفاده نمودند. آن‌ها در مطالعه خود با استفاده از روش PCA، از میان داده‌های ساعت آفتابی، تندی باد، بارش روزانه، رطوبت نسبی، فشار مطلق و تبخیر روزانه اقدام به انتخاب بهترین متغیرهای ورودی به ANN نمودند (۱۹). مطالعه نجفی و همکاران (۲۵) از روش ICA برای گزینش بهترین متغیرهای ورودی ریزمقیاس نمایی بارش روزانه با استفاده از مدل ANFIS و SVM در حوضه ویلامت^۶ بهره گرفتند (۲۵).

بررسی منابع نشان می‌دهد که اکثر مطالعات صرفاً به کاربرد روش‌های مختلف انتخاب متغیر اکتفا نموده‌اند و پژوهش‌های کمی در خصوص مقایسه اثرات این روش‌ها در عملکرد ریزمقیاس نمایی انجام شده است. مطالعه یانگ و همکاران (۳۷) از جمله این معدود مطالعات به شمار می‌رود. این مطالعه با هدف ارزیابی عملکرد روش‌های SRA^۱، CA^۱، ParCA^۷ در ریزمقیاس نمایی بارش روزانه در نواحی شمالی کشور چین انجام شده است. در این مطالعه وی با استفاده از آزمون‌های مقایسه‌ای متعدد اثرات هر کدام از روش‌های انتخاب متغیر را در بهبود نتایج مورد ارزیابی قرار داده است. نتایج مطالعه آن‌ها نشان داد که روش SRA در ریزمقیاس نمایی بارش اثرات مطلوب‌تری نسبت به دو روش دیگر دارد (۳۷). همچنین مطالعه همای و همکاران (۱۱) نیز با هدف مقایسه عملکرد روش‌های SRA

مواد و روش‌ها

معرفی منطقه مورد مطالعه و توصیف داده‌ها

شهرستان بیرجند مرکز استان خراسان جنوبی در شمال رشته کوه‌های باقران و در محدوده‌ی ۵۹ درجه و ۱۲ دقیقه و صفر ثانیه طول جغرافیایی و ۳۲ درجه و ۵۲ دقیقه و صفر ثانیه عرض جغرافیایی قرار گرفته است. میانگین کمینه و بیشینه دمای سالانه بیرجند به ترتیب ۸ و ۲۴ درجه سانتی گراد، بارش سالانه ۱۲۰ میلی‌متر و میانگین تبخیر و تعرق پتانسیل ۲۷۰۰ میلی‌متر است. بر اساس شاخص‌های اقلیم دومارتن و ایوانف، شهر بیرجند دارای اقلیم خشک- کویری می‌باشد (۱۸).

دشت اردبیل مشرف بر ارتفاعات بخش غربی رشته کوه البرز (ارتفاعات تالش)؛ در محدوده ۴۸ درجه و ۱۷ دقیقه و صفر ثانیه طول جغرافیایی و ۳۸ درجه و ۱۵ دقیقه و صفر ثانیه عرض جغرافیایی و در امتداد دامنه شرقی سبلان واقع شده است. مساحت آن حدود ۹۹۰ کیلومتر مربع و متوسط بارش سالانه در ایستگاه سینوپتیک اردبیل حدود ۳۵۰ میلی‌متر می‌باشد بیشترین و کمترین بارش به ترتیب در ماه‌های اردیبهشت و مرداد اتفاق می‌افتد. میانگین دمای سالانه ایستگاه سینوپتیک اردبیل ۷ درجه سانتی گراد است و یکی از نواحی سردسیر ایران محسوب می‌شود. براساس طبقه‌بندی اقلیمی دومارتن و ایوانف، شهر اردبیل دارای اقلیم سرد- نیمه‌خشک می‌باشد (۳۲).

در پژوهش‌های تغییر اقلیم دو نوع داده در نظر گرفته می‌شود: داده‌های مشاهداتی- پیش‌بینی شونده و داده‌های مدل‌های GCMs -

- 1-Genetic Algorithm
- 2-Nash Sutcliff Efficiency
- 3-Mallow Coefficient
- 4-Principal Component Analysis
- 5-Independent Component Analysis-
- 6-Willamette
- 7-Partial Correlation Analysis

- 8-Least Absolute Shrinkage And Selection Operator-LASSO
- 9-Québec

ورودی به نحوی صورت پذیرد تا بین پیش‌بینی‌کننده‌ها و متغیر مورد نظر (پیش‌بینی‌شونده) همبستگی قوی وجود داشته و از حیث مشخصات فیزیکی ارتباط معنی‌داری داشته باشند. همچنین بهتر است مجموعه پیش‌بینی‌کننده‌ها شامل متغیرهایی باشند که توصیف کاملی از رطوبت، گردش هوا و ارتباط بین لایه‌های مختلف جو بالای زمین را دارا باشند (۳۳). همان‌طور که قبلاً توضیح داده شد، در این مطالعه با استفاده از خروجی مدل CanESM2، تعداد ۲۶ متغیر بزرگ مقیاس به عنوان پیش‌بینی‌کننده انتخاب شدند. این ۲۶ متغیر که در (جدول ۱) ارایه شده‌اند، به صورت روزانه و در سطوح مختلف زمین موجود می‌باشند. با توجه به اینکه هدف اصلی مطالعه پیش‌رو، ارزیابی اثرات روش‌های مختلف انتخاب متغیر ورودی در ریزمقیاس نمایی آماری بارش روزانه می‌باشد، در ادامه توضیحات مربوط به ۴ روش SRA، PCA، CA و ParCA ارایه می‌شود.

روش SRA

این روش می‌تواند به دو صورت اضافه کردن متغیرهای وابسته (پیش‌رو) و یا حذف آن‌ها (پس‌رو) از معادله رگرسیون انجام شود. برای شروع ابتدا یک مدل بر متغیر مستقل برازش داده می‌شود. در ادامه برای انتخاب متغیرهای ورودی به مدل لازم است تا یک فرآیند تکراری انجام شود. تمامی متغیرها لازم است تا یک مدل بر متغیر وابسته برازش دهند و مقدار آماره $\rho - value$ برای ضریب رگرسیونی تخمین زده توسط آن‌ها تعیین گردد. در اولین گام متغیری وارد مدل خواهد شد که مقدار $\rho - value$ آن از حد ورودی کمتر بوده و ضمناً کمترین مقدار $\rho - value$ را در بین متغیرهای مستقل دارا باشد. در گام‌های بعدی نیز نحوه ورود متغیرها نیز به همین شکل خواهد بود. لکن در هر گام چنانچه مقدار $\rho - value$ متغیری از حد خروجی بیشتر باشد از مدل خارج خواهد شد (۷). همچنین در هر گام مقدار ضریب اختصاص داده شده به هر متغیر، و مقدار خطای برآورد مشخص می‌شود.

روش PCA

رویکرد PCA نوعی از تجزیه و تحلیل آماری است که تعداد کمتری از مجموعه اولیه را به نام مؤلفه‌های اصلی از میان متغیرهای اولیه گزینش می‌کند، به طوری که بیشترین (ضروری‌ترین) اطلاعات مجموعه اولیه حفظ شده و از مابقی اطلاعات غیر ضروری صرفه‌نظر می‌شود. در این روش با استفاده از n متغیر $x_1, x_2, x_3, \dots, x_n$ مجموعه اولیه‌ی داده‌ها (X) تعداد p متغیر مستقل به صورت $PC_1, PC_2, PC_3, \dots, PC_p$ ایجاد می‌شود. رابطه (۱) چگونگی تشکیل متغیرهای مستقل را نشان می‌دهد:

$$PC_j = \sum_{i=1}^n W_{i,j} x_i \rightarrow j = 1 : p \quad (1)$$

پیش‌بینی‌کننده. متغیرهای پیش‌بینی‌شونده همان مؤلفه‌های اقلیم می‌باشند که در ایستگاه‌های هواشناسی و در مقیاس محلی ثبت می‌شوند. داده‌های پیش‌بینی‌کننده، متغیرهای مستقلی می‌باشند که توصیف‌کننده وضعیت رطوبت، چرخش هوا و دمای اتمسفر در سطوح مختلف ارتفاعی می‌باشند. مقیاس داده‌های پیش‌بینی‌کننده بسیار بزرگ بوده و عموماً به آن‌ها داده‌های بزرگ مقیاس اطلاق می‌شود. همچنین باید توجه نمود که به‌منظور ارزیابی توانمندی مدل‌های GCMs، خروجی این مدل‌ها همواره برای ۲ دوره تاریخی و آینده ارایه می‌شود. دوره تاریخی شامل مقادیر متغیرهای پیش‌بینی در ۳۰ سال گذشته می‌باشد. طبق توصیه IPCC دوره تاریخی بین سال‌های ۱۹۶۰ تا ۱۹۹۰ می‌باشد. اما به دلیل اینکه در این بازه زمانی ممکن است که برخی از ایستگاه‌های هواشناسی فعال نبوده و ثبت داده صورت نمی‌گرفته است، لذا خروجی مدل‌های GCMs تا سال ۲۰۰۵ برای دوره تاریخی در نظر گرفته می‌شود. دوره آینده نیز از ابتدای سال ۲۰۰۶ تا انتهای سال ۲۰۹۹ می‌باشد. در این پژوهش ابتدا داده‌های مشاهداتی بارش روزانه حداقل سال‌های ۱۹۷۷ تا انتهای سال ۲۰۰۴ در ۲ ایستگاه سینوپتیک بیرجند و اردبیل به عنوان دوره تاریخی و مدل CanESM2 (Canadian Earth System Model) به عنوان مدل GCMs در نظر گرفته شد. این مدل که محصول مرکز اقلیم‌شناسی Canadian Climate Data and Scenarios (www.ccds-dscc.ec.gc.ca) می‌باشد، کل سطح زمین را به‌منظور شبیه‌سازی مقادیر اقلیمی به ۱۲۸ گره افقی و ۶۴ گره عمودی تقسیم نموده است. از این رو فاصله افقی و عمودی هر گره حدوداً ۲/۸۱۲۵ درجه می‌باشد. در حال حاضر این پایگاه در کنار مقادیر روزانه مدل CanESM2، شبیه‌سازی‌های ماهانه کلیه مدل‌های GCMs را برای بارندگی، متوسط درجه حرارت، سرعت باد و غیره در اختیار کاربران قرار می‌دهد. بدین ترتیب متغیرهای پیش‌بینی‌کننده‌ی بزرگ مقیاس برای حداقل سال‌های ۱۹۷۷ الی ۲۰۰۴ میلادی از این پایگاه استخراج گردید و سال‌های ۱۹۹۵، ۱۹۹۶ الی ۲۰۰۰ و ۲۰۰۱ الی ۲۰۰۴ میلادی به ترتیب به عنوان دوره آموزش، صحت‌سنجی و آزمون فرض شد.

انتخاب پیش‌بینی‌کننده‌ها

در ریزمقیاس نمایی آماری به روش توابع انتقال، بین متغیرهای بزرگ مقیاس پیش‌بینی‌کننده و مؤلفه‌های اقلیمی مشاهداتی یک ارتباط ایجاد می‌شود. اینکه تمام و یا بخشی از متغیرهای پیش‌بینی‌کننده برای ایجاد ارتباط با داده‌های مشاهداتی استفاده شود، تحت عنوان انتخاب پیش‌بینی‌کننده‌ها و یا متغیرهای ورودی شناخته می‌شود. انتخاب پیش‌بینی‌کننده‌ها اولین گام ریزمقیاس نمایی آماری محسوب می‌شود. در حالت کلی باید اهتمام داشت انتخاب متغیرهای

جدول ۱- جزئیات متغیرهای پیش‌بینی کننده
Table 1- Details of predictor variables utilized

ارتفاع Level (hPa)	چرخش Circulation	رطوبت Humidity	دما Temperature
سطح زمین Surface	X1- میانگین فشار سطح دریا Mean sea level pressure (mslp)	X22- رطوبت نسبی Near surface relative humidity (rhum)	X26- دمای میانگین Mean temperature (temp)
	X2- سرعت جریان هوای زمین گرد Geostrophic air flow velocity (p1_f)	X25- رطوبت ویژه نزدیک سطح Near surface specific humidity (shum)	
	X3- سرعت نصف‌النهاری-عرض Meridional velocity component (p1_v)		
	X4- سرعت ناحیه‌ای-طول Zonal velocity component (p1_u)		
	X5- چرخش هوا Vorticity (p1_z)		
	X6- جهت باد Wind direction (p1th)		
	X7- دیوژانس Divergence (p1zh)		
500 hPa	X8- سرعت جریان هوای زمین گرد Geostrophic air flow velocity (p5_f)	X23- رطوبت ویژه Specific humidity (s500)	
	X9- سرعت نصف‌النهاری-عرض Meridional velocity component (p5_v)		
	X10- سرعت ناحیه‌ای-طول Zonal velocity component (p5_u)		
	X11- چرخش هوا Vorticity (p5_z)		
	X12- جهت باد Wind direction (p5th)		
	X13- دیوژانس Divergence (p5zh)		
	X20- ارتفاع ژئوپتانسیل Geopotential height (p500)		
850 hPa	X14- سرعت جریان هوای زمین گرد Geostrophic air flow velocity (p8_f)	X24- رطوبت ویژه Specific humidity (s850)	
	X15- سرعت نصف‌النهاری-عرض Meridional velocity component (p8_v)		
	X16- سرعت ناحیه‌ای-طول Zonal velocity component (p8_u)		
	X17- چرخش هوا Vorticity (p8_z)		
	X18- جهت باد Wind direction (p8th)		
	X19- دیوژانس Divergence (p8zh)		
	X21- ارتفاع ژئوپتانسیل Geopotential height (p850)		

*انتخاب شده از (۲۲).

تغییر پذیری) مربوط به مجموعه اولیه را در برداشته باشد و مولفه‌ی اصلی دوم (PC_2) نیز بیشترین واریانس در نظر گرفته نشده توسط مولفه اول را شامل شود. این روند تا انتها برای کلیه مولفه‌های اصلی

در رابطه (۱)، x_i : معرف اُمین متغیر اولیه و W_{ij} : ضریب اُمین متغیر اولیه در اُمین مؤلفه اصلی (PC_j) می‌باشد. ضرایب W به نحوی تعیین می‌شوند که اولین مولفه‌ی اصلی (PC_1) بیشترین واریانس

ریزمقیاس نمایی با استفاده از شبکه عصبی مصنوعی

معرفی شبکه عصبی توسط توسط مک کولاک و پیتس در سال ۱۹۴۳ انجام شد (۲۳). اما توسعه این روش و کاربردهای گسترده آن از ابتدای سال ۱۹۸۶ توسط راملهارت و همکاران انجام شد (۲۹). شبکه‌های عصبی با اقتباس از عملکرد مغز انسان پیاده‌سازی شده‌اند. در این شبکه‌ها، نرون‌ها به عنوان دریافت کننده اطلاعات عمل می‌کنند. هر نرون به عنوان کوچکترین واحد عملیاتی شبکه عصبی، با استفاده از ضرایبی که به هر کدام از اطلاعات ورودی اختصاص می‌دهد بخشی از اطلاعات ورودی به شبکه عصبی را دریافت نموده و مجموعه نرون‌ها در کنار یکدیگر، یک تحلیل از واقعیت پیرامون خود را ایجاد می‌کنند. این تحلیل از اطلاعات منجر به یک خروجی (داده‌های مشاهداتی) می‌شود و عموماً با یک خطا همراه می‌باشد که از آن تحت عنوان خطای شبیه‌سازی شبکه عصبی یاد می‌شود. اطلاعات ورودی، نرون‌ها و خروجی هر کدام در یک لایه قرار دارند. یک شبکه عصبی یا می‌تواند به صورت ساده شامل یک لایه برای ورود اطلاعات، یک لایه برای نرون‌ها و یک لایه برای شبیه‌سازی خروجی باشد؛ و یا اینکه به صورت ترکیبی از چندین لایه شبکه عصبی تنظیم شود (۱۰).

در صورتی که یک شبکه عصبی به صورت ترکیبی از چندین لایه باشد، تخمین لایه اول با یک تابع انتقال به عنوان ورودی لایه دوم اعمال می‌شود. نرون‌های موجود در لایه دوم اطلاعات ورودی به این لایه را همراه با خطای مشخصی تخمین خواهند زد. خروجی لایه دوم نیز از طریق یک تابع انتقال و به عنوان ورودی به لایه بعدی معرفی می‌گردد. بدین ترتیب به تعداد لایه‌های موجود در شبکه عصبی، اطلاعات ورودی به شبکه عصبی مورد تحلیل واقع می‌شود و در آخرین لایه شبیه‌سازی شبکه عصبی با خروجی واقعی مقایسه می‌شود (۲۰). عموماً در مطالعات مربوط به کاربرد شبکه عصبی لایه‌های میانی شبکه عصبی را لایه‌های پنهان نام‌گذاری می‌کنند و مجموعه لایه‌های پنهان، نرون‌ها و توابع انتقال و آموزش را تحت عنوان معماری شبکه عصبی بیان می‌کنند. لازم به ذکر است که مطالعات بسیار زیادی وجود دارد که جهت رسیدن به جواب‌های بهتر در شبکه عصبی، از معماری‌های سنگین استفاده نموده‌اند. درعین حال مطالعات زیادی نیز وجود دارند که برای رسیدن به یک جواب مطلوب در شبکه‌های عصبی از معماری‌های سبک استفاده نموده‌اند (۱). آنچه برای عملکرد یک شبکه عصبی حایز اهمیت است، تخمین دقیقی خروجی می‌باشد. تخمین شبکه عصبی تابعی از دو مقوله می‌باشد: نخست اطلاعاتی است که از داده‌های ورودی دریافت نموده و دیگر تحلیل دقیق نرون‌ها به نحوی که خطای شبیه‌سازی حداقل شود. تخمین شبکه عصبی به طور کاملاً مستقیم به ضرایب نرون‌ها در هر لایه مربوط می‌شود. به منظور بهینه‌یابی این ضرایب از یک تابع

در نظر گرفته می‌شود. به منظور تعیین ضرایب هر کدام از مؤلفه‌های اصلی مراحل زیر انجام شد:

۱- استانداردسازی داده‌ها به نحوی که میانگین و انحراف معیار داده‌ها به ترتیب به صفر و یک متمایل شود. ۲- ماتریس همبستگی (کوواریانس): میزان همبستگی بین متغیرهای اولیه در این ماتریس مشخص می‌شود. با تعیین مقادیر ویژه این ماتریس می‌توان ضرایب مؤلفه‌های اصلی را برآورد نمود. ۳- در نهایت با بررسی این که هر کدام از مؤلفه‌های اصلی چه میزان از واریانس مجموعه اصلی را دربر گرفته است، می‌توان تعداد مؤلفه‌های اصلی را مشخص نمود. در این مطالعه حد ۹۰ درصد از واریانس متغیرهای پیش‌بینی کننده به عنوان آستانه انتخاب تعداد مؤلفه‌های اصلی انتخاب شد.

CA

این آزمون براساس ضریب همبستگی بین داده‌های بزرگ مقیاس و مشاهداتی انجام می‌شود. در این پژوهش بر اساس ضریب همبستگی پیرسون (Pearson) میزان همبستگی بین متغیرها و بارندگی مشاهداتی اندازه‌گیری می‌شود. فرض صفر مبنی بر عدم وجود همبستگی می‌باشد و سطح معنی‌داری ۹۵ درصد مورد نظر می‌باشد. رابطه (۲) ضریب همبستگی پیرسون را نشان می‌دهد (۲۷):

$$R_{xy} = \frac{1}{n-1} \sum_{i=1}^n \left[\left(\frac{x_i - \mu_x}{\sigma_x} \right) \cdot \left(\frac{y_i - \mu_y}{\sigma_y} \right) \right] \quad (2)$$

در رابطه (۲) x : متغیر پیش‌بینی کننده و y : بارندگی مشاهداتی، n : تعداد داده‌ها، σ : انحراف از معیار و μ : میانگین داده‌ها می‌باشد. همچنین منظور از Cov کوواریانس بین ۲ متغیر x و y می‌باشد.

روش ParCA

در این روش انتخاب متغیر بر اساس ضرایب همبستگی جزئی بیان می‌شود. این آزمون همبستگی بین دو متغیر را با برداشتن اثر سایر متغیرها، بیان می‌کند. سطح معنی‌داری و کیفیت فرض صفر مشابه آزمون همبستگی می‌باشد. همبستگی بین دو متغیر پیش‌بینی کننده (y) و بارندگی مشاهداتی (x) درحالی که اثر سایر متغیرهای پیش‌بینی کننده برداشته شده باشد، طبق رابطه (۳) بیان می‌شود:

$$R_{xy,z} = \frac{R_{xy} - R_{xz} \cdot R_{yz}}{\sqrt{(1 - R_{xz}^2) \cdot (1 - R_{yz}^2)}} \quad (3)$$

در رابطه فوق R ضریب همبستگی پیرسون می‌باشد. در این مطالعه از میان متغیرهایی که همبستگی معنی‌داری با بارش روزانه داشته باشند، متغیرهایی که بیشترین ضریب همبستگی و همبستگی جزئی را دارا باشند، انتخاب شدند. در واقع متغیرهایی انتخاب می‌شوند که دارای بیشترین معنی‌داری با مقادیر بارش روزانه می‌باشند.

متعددی استفاده شد. در ابتدا میزان دقت ریزمقیاس نمایی تحت تأثیر هر کدام از روش‌های انتخاب متغیر با شاخص‌های ارزیابی مورد بررسی قرار گرفت. این شاخص‌ها که از جمله پرکاربردترین شاخص‌های ارزیابی در مطالعات تغییر اقلیم می‌باشند در روابط (۴) تا (۶) ارائه شده‌اند:

$$RMSE = \sqrt{\frac{\sum_i^n (X_i - Y_i)^2}{n}} \quad (4)$$

$$R = \frac{\left[\sum_i^n (X_i - \bar{X}) \times (Y_i - \bar{Y}) \right]}{\sum_i^n (X_i - \bar{X}) \times \sum_i^n (Y_i - \bar{Y})} \quad (5)$$

$$NASH = 1 - \frac{\sum_i^n (X_i - Y_i)^2}{\sum_i^n (X_i - \bar{X})^2} \quad (6)$$

در روابط فوق X و Y به ترتیب مقادیر بارندگی روزانه مشاهداتی و ریزمقیاس شده، \bar{X} و \bar{Y} به ترتیب میانگین این مقادیر و n تعداد داده‌های باشد. پژوهش‌های هم‌امی و همکاران (۱۱)، بانگ و همکاران (۳۷)، هارفام و ویلیام (۱۲) و ناصری و همکاران (۲۶) از جمله پژوهش‌هایی می‌باشند که در مطالعه خود از این شاخص‌ها بهره گرفته‌اند. بعد از محاسبه شاخص‌های ارزیابی از چند آزمون مقایسه‌ای جهت تعیین برتری روش‌های مختلف انتخاب متغیر استفاده شد. مقایسه مشخصه‌های آماری، مقایسه نمودار چگالی و توزیع فراوانی با استفاده از Violin Plot و بررسی توانمندی روش‌ها در تشخیص سری‌های تر و خشک با استفاده از Contingency Table Event از جمله این آزمون‌های باشد.

لازم به ذکر است که جهت تخمین روزهای تر و خشک با استفاده از Contingency Table Event، حد آستانه روز تر و خشک معادل ۱ میلی‌متر در روز در نظر گرفته شد (۲۸). در واقع این جدول مقایسه روزهای تر و خشک مشاهداتی و شبیه‌سازی را به صورت متقابل نشان می‌دهد (۳). به منظور کمی کردن دقت روش‌های مختلف انتخاب متغیر از شاخص CSI^4 استفاده شد (۲۸):

$$CSI = \frac{Hit}{Hit + Miss + False\ alarm} \quad (7)$$

در رابطه فوق، Hit: تعداد تشخیص صحیح روز تر، Miss: تشخیص اشتباه در روز تر و False alarm: تشخیص اشتباه در روز خشک می‌باشد. این شاخص نسبت تعداد تشخیص‌های درست روزهای تر را نشان می‌دهد و چنانچه روزهای تر به درستی تشخیص داده شوند، روزهای خشک مشاهداتی نیز به درستی شبیه‌سازی خواهند شد. در واقع اگر تعداد تشخیص‌های اشتباه به سمت صفر میل

آموزش استفاده می‌شود. توابع آموزش ضرایب نرون‌ها را در یک بهینه‌سازی خطی به نحوی تعیین می‌کند تا شبکه عصبی دارای کمترین خطای تخمین شود. در کنار توابع آموزش، یکی دیگر از مهم‌ترین اقداماتی که سبب توسعه شبکه عصبی شده است، کاربست الگوریتم پس انتشار خطا^۱ BPA می‌باشد. الگوریتم BPA بدین مفهوم می‌باشد که شبکه عصبی پس از یک بار تخمین مقدار خروجی، خطای شبیه‌سازی را بهبود بخشد. فرآیند بهبود خطای شبیه‌سازی از طریق انتشار خطا در لایه‌های مختلف شبکه عصبی می‌باشد. در این روش درصدهای مختلف خطای شبیه‌سازی به نحوی در لایه‌های مختلف پخش (انتشار) می‌شود که مقدار کل خطا در شبیه‌سازی بعدی به حداقل برسد. اگر فرآیند انتشار خطا از لایه نخست به سمت انتهایی‌ترین لایه به سمت اول صورت پذیرد به آن پس‌رو اطلاق می‌شود (۲۹).

در این مطالعه معماری شبکه عصبی بدین صورت پیاده‌سازی شد که: داده‌های بزرگ مقیاس انتخاب شده توسط هر کدام از روش‌های انتخاب متغیر به عنوان ورودی، بارندگی روزانه به عنوان خروجی، تعداد سه لایه‌ی پنهان، تعداد سه نرون در هر لایه‌ی پنهان، تابع لونیگ‌مارکوارت^۲ جهت آموزش نرون‌ها در تخمین ضرایب، تابع محرکه تانژانت سیگموئید^۳ به عنوان تابع انتقال لایه‌های پنهان، تابع محرکه خطی به عنوان تابع انتقال در لایه خروجی (آخرین لایه) و تعداد تکرارهای شبیه‌سازی شبکه عصبی ۵۰۰ مرتبه در نظر گرفته شد. همچنین به منظور جلوگیری از اتلاف وقت و با توجه به احتمال قرارگیری الگوریتم آموزش در بهینه‌های محلی تعیین مقادیر بهینه ضرایب، سعی شد تا از شرایط خاتمه استفاده شود. شرط خاتمه در این مطالعه تعداد ۲۰ مرتبه عدم بهبود در تکرارهای شبکه عصبی می‌باشد. بدین ترتیب چنانچه الگوریتم شبکه عصبی در ۲۰ تکرار متوالی بهبود خاصی در عملکرد شبکه عصبی نداشت، خاتمه خواهد یافت. همچنین انتخاب معماری سه لایه پنهان همراه با ۳ نرون بدین جهت می‌باشد که با سعی و خطاهای مختلف در اجرای کد شبکه عصبی مشخص شد که افزایش تعداد لایه‌های پنهان و تعداد نرون‌ها، موجب بهبود معنی‌دار عملکرد شبکه عصبی نمی‌شود.

ارزیابی عملکرد

پس از مشخص شدن متغیرهای ورودی به شبکه عصبی جهت ریزمقیاس نمایی کردن بارش روزانه، به منظور تشخیص تأثیر هر کدام از روش‌های انتخاب متغیر بر فرآیند ریزمقیاس نمایی از آزمون‌های

1-Back Propagation Algorithm
2-Levenberg-Marquardt
3-Hyperbolic tangent sigmoid

4-Critical Success Index

کند، شاخص CSI به بیشترین مقدار خود خواهد رسید.

نتایج و بحث

در ابتدا به بررسی متغیرهای انتخاب شده در هر کدام از روش‌های انتخاب متغیر پرداخته می‌شود. طبق روش SRA، تعداد ۱۳ متغیر در ایستگاه بیرجند و تعداد ۸ متغیر در ایستگاه اردبیل به عنوان ورودی مدل انتخاب شدند. ضرایب هر کدام از متغیرها، عرض از مبدأ معادله رگرسیونی و مقادیر خطای این روش در (جدول ۲) ارائه شده است. با استناد به مقدار خطای RMSE مشخص می‌شود که معادله رگرسیون در ایستگاه بیرجند با اتکا به متغیرهای انتخاب شده دارای دقت بهتری نسبت به ایستگاه اردبیل می‌باشد. همچنین به لحاظ تعداد متغیر، معادله رگرسیونی ایستگاه اردبیل دارای متغیرهای کمتری نسبت به ایستگاه بیرجند می‌باشد.

در مطالعه خاشعی و همکاران (۲۱) نیز مقدار جذر مربعات خطای معادله رگرسیونی متغیرهای انتخاب شده پس از ۱۸ گام پیاپی با ۱۷ متغیر انتخاب شده به ۱/۸۳ میلی‌متر رسید (۲۱). در روش PCA تعداد ۲۶ متغیر پیش‌بینی کننده به ترتیب به ۱۳ و ۱۶ مؤلفه اصلی در ایستگاه بیرجند و اردبیل کاهش پیدا کرد. نمودار ScreePlot به طور هم‌زمان درصد واریانس هر کدام از مؤلفه‌های اصلی، مقدار واریانس پوشش داده شده توسط هر کدام و درصد تجمعی واریانس را نشان

می‌دهد. با استناد (شکل ۱) مشخص می‌شود که چنانچه ۹۰ درصد از واریانس متغیرهای پیش‌بینی کننده مورد نظر باشد، در ایستگاه اردبیل ۱۶ مؤلفه‌ی اول و در ایستگاه بیرجند ۱۳ مؤلفه اول این میزان از تغییرپذیری مجموعه اول را به طور کامل پوشش می‌دهند (قسمت هاشور خورده‌ی شکل ۱). البته باید توجه داشت که اولین مؤلفه‌ی اصلی در بیرجند و اردبیل به ترتیب حدود ۵۰ و ۴۶/۵ درصد از کل واریانس مجموعه اول را دربر گرفته است.

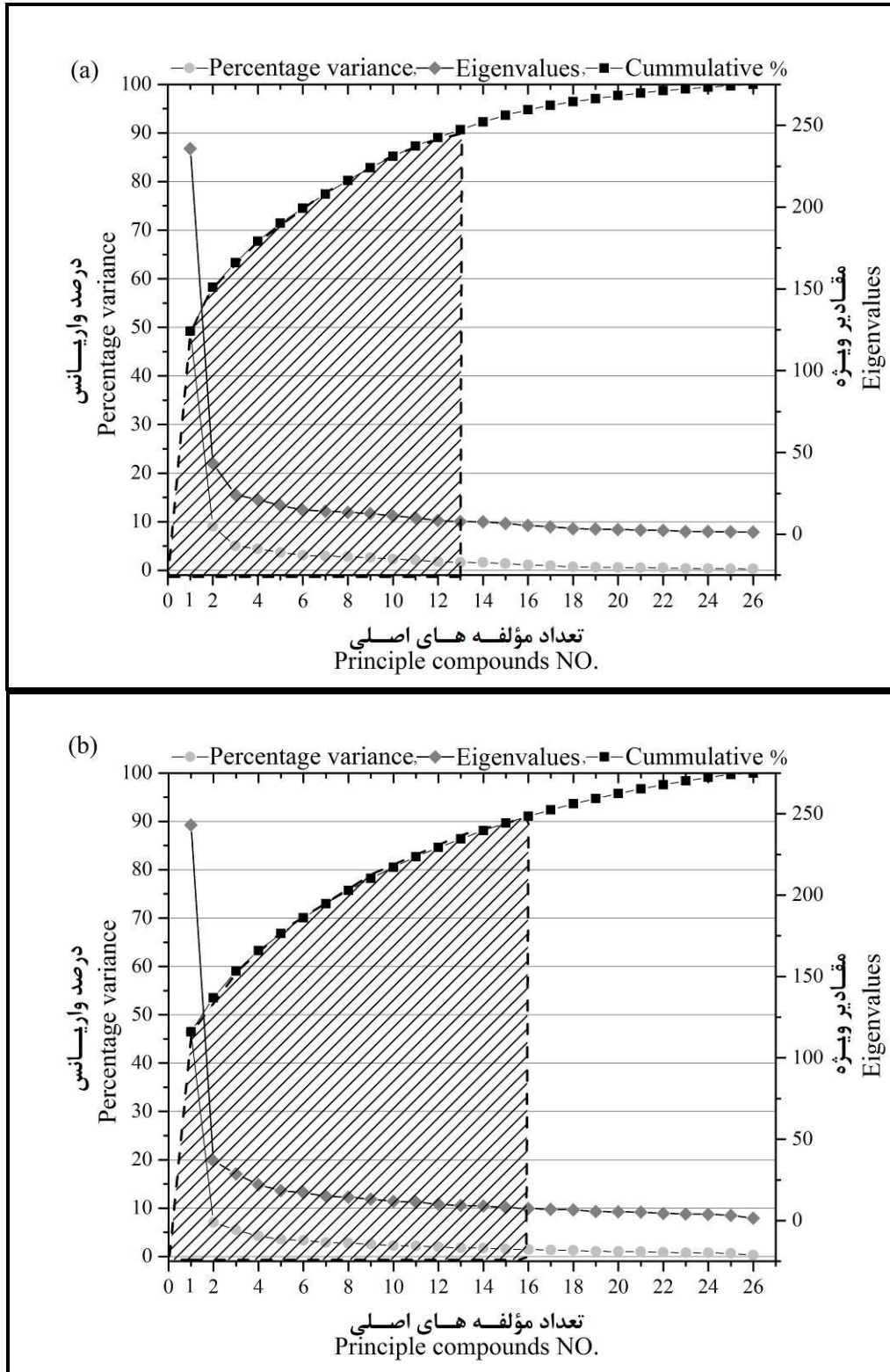
در انتخاب پیش‌بینی کننده به روش CA، بر اساس مقدار سطح معنی‌داری ۰/۰۵ تعداد متغیرهای ورودی در هر ایستگاه همراه با ضریب همبستگی پیرسون در (جدول ۳) مشخص شده است. همان‌طور که از اعداد مندرج در (جدول ۳) مشخص می‌شود، همبستگی متغیرهای پیش‌بینی کننده در ایستگاه اردبیل با مقادیر بارش روزانه کمتر از ۰/۱ و بسیار ضعیف می‌باشد. این درحالی است که در ایستگاه بیرجند ضریب همبستگی بین بارش روزانه و متغیرهای رطوبت و دیوژانس بسیار بهتر و به ترتیب ۰/۵۴ و ۰/۲۷- می‌باشد.

در نهایت متغیرهایی که به روش ParCA برای ریزمقیاس‌نمایی انتخاب شده‌اند در (جدول ۴) نمایش داده شده‌اند. آزمون ParCA به لحاظ عملکرد شبیه آزمون همبستگی می‌باشد؛ با این تفاوت که برای محاسبه ضریب همبستگی اثر سایر مؤلفه‌ها را خنثی می‌کند. به همین جهت انتظار می‌رود که در این آزمون تعداد متغیر کمتری شرایط سطح معنی‌داری را داشته باشد.

جدول ۲- متغیرهای پیش‌بینی کننده انتخاب شده در روش SRA

Table 2- Selected predictor variables in SRA

ایستگاه بیرجند		ایستگاه اردبیل	
Birjand		Ardebil	
متغیر	ضریب	متغیر	ضریب
Variable	Coefficient	Variable	Coefficient
X6	-0.0005	X1	-0.7175
X7	0.0778	X3	-0.059
X8	0.301	X5	-0.1529
X9	-0.3801	X15	-0.1178
X10	0.2754	X17	0.0736
X12	-0.0012	X21	0.5173
X13	-0.0532	X22	0.0595
X17	0.1088	X26	-0.6069
X18	-0.0004		
X20	-0.2421		
X22	0.8435		
X23	0.1195		
X25	0.0469		
مقدار عرض از مبدأ	0.4	مقدار عرض از مبدأ	0.76
RMSE (mm)	1.4309	RMSE (mm)	2.3484



شکل ۱- نمودار Scree Plot مقادیر ویژه در مقابل مؤلفه‌های اصلی در کنار درصد واریانس در مقابل مؤلفه‌های اصلی ایستگاه بیرجند (a) و ایستگاه اردبیل (b)

Figure 1- Scree plot of Eigenvalues vs. components along with percentage variances vs. components for Birjand (a) & Ardebil (b)

جدول ۳- متغیرهای دارای همبستگی معنی‌دار

Table 3- Predictors with significant correlation

ایستگاه اردبیل Ardebil station		ایستگاه بیرجند Birjand station	
متغیر Variable	ضریب همبستگی Correlation Coefficient	متغیر Variable	ضریب همبستگی Correlation Coefficient
X20	-0.076	X10	0.0324
X25	-0.074	X19	-0.277
X26	-0.091	X22	0.546

جدول ۴- متغیرهای دارای همبستگی جزئی معنی‌دار

Table 4- Predictors with significant partial correlation

ایستگاه اردبیل Ardebil station		ایستگاه بیرجند Birjand station	
متغیر Variable	ضریب همبستگی جزئی Partial Correlation Coefficient	متغیر Variable	ضریب همبستگی جزئی Partial Correlation Coefficient
X1	-0.041	X9	-0.088
X5	-0.042	X10	0.088
X26	-0.062	X22	0.373

مندرج در (جدول ۵) مشخص می‌شود که در ایستگاه بیرجند روش‌های CA دارای عملکرد بهتری نسبت به SRA می‌باشد. به عبارت دیگر به کارگیری روش‌های مبتنی بر همبستگی باعث ارتقای بیشتر فرآیند ریزمقیاس نمایی بارش در مناطق خشک- کویری خواهد شد. لکن در مناطقی نظیر اردبیل (سرد- نیمه‌خشک)، ریزمقیاس نمایی بر اساس روش‌های انتخاب متغیر مبتنی بر رگرسیون توصیه می‌شود. این موضوع با یافته‌های پژوهش یانگ (۳۷) کاملاً مطابقت دارد. نتایج مطالعه وی که به منظور بررسی اثر روش‌های مختلف انتخاب متغیر در ریزمقیاس نمایی بارش در شمال چین انجام شده است، نشان می‌دهد که ریزمقیاس نمایی با استفاده از روش‌های انتخاب متغیر مبتنی بر رگرسیون دارای عملکرد بهتری نسبت به روش‌های مبتنی بر همبستگی می‌باشد (۳۷).

نمودار Violin در واقع نشان دهنده شکل توزیع آماری یک نمونه و پراکنش داده‌ها می‌باشد. (شکل ۳) نمودار Violin بارندگی مشاهداتی و ریزمقیاس شده را همراه با مقدار میانگین (دایره قرمز رنگ) برای دو ایستگاه بیرجند و اردبیل نشان می‌دهد. مقایسه شکل توزیع بارندگی مشاهداتی و ریزمقیاس شده (شکل ۳) نشان می‌دهد که پراکنش مقادیر بارش روزانه در تمام روش‌های انتخاب متغیر در ایستگاه اردبیل زیاد مطلوب نمی‌باشد. این در حالی است که شکل توزیع‌های آماری مشاهداتی و ریزمقیاس شده در ایستگاه بیرجند از تشابه بیشتری برخوردار می‌باشد.

(جدول ۶ و ۷) مقایسه رخدادهای تر و خشک مشاهداتی و ریزمقیاس شده را برای ایستگاه‌های بیرجند و اردبیل نشان می‌دهند. با مقایسه تعداد روزهای تر و خشک مشاهداتی ایستگاه بیرجند در

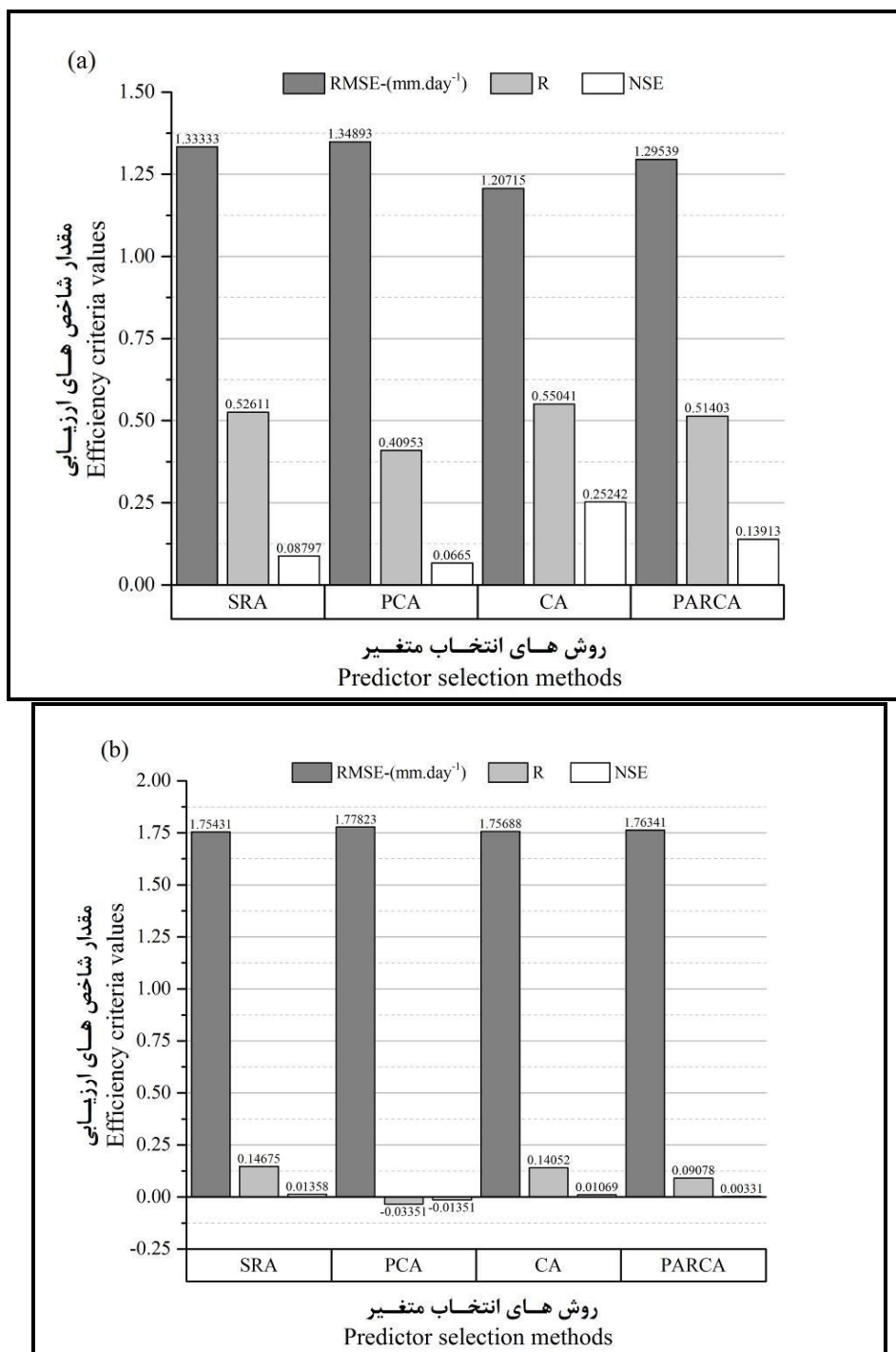
با در نظر گرفتن نتایج روش‌های انتخاب پیش‌بینی کننده مشخص می‌شود که در ایستگاه بیرجند متغیرهای سرعت ناحیه‌ای و رطوبت نسبی و در ایستگاه اردبیل متغیرهای میانگین فشار سطح دریا، چرخش هوا و دمای میانگین جزء متغیرهای غالب و ورودی به شمار می‌روند. از این حیث نتایج پژوهش پیش‌رو با نتایج مطالعه ناصری و همکاران (۲۶) کاملاً یکسان می‌باشد. نتایج مطالعه ناصری و همکاران (۲۶) که از روش ParCA جهت انتخاب متغیرهای ورودی ریزمقیاس نمایی استفاده شده بود، نشان داد که متغیرهای رطوبت نسبی و سرعت ناحیه‌ای در سطوح مختلف فشار به عنوان متغیر ورودی برگزیده شدند (۲۶).

در ادامه نتایج مربوط به آزمون‌های مقایسه‌ای ارائه می‌گردد. ابتدا شاخص‌های ارزیابی مربوط به تأثیر هر کدام از روش‌ها در عملکرد شبکه عصبی در ریزمقیاس نمایی و مربوط به دوره آزمون (۲۰۰۴-۲۰۰۱) در (شکل ۲) ارایه شده است. با استناد به مقادیر شاخص‌های ارزیابی در دو ایستگاه می‌توان بیان نمود که ریزمقیاس نمایی بارش روزانه با استفاده از روش CA دارای عملکرد بهتری نسبت به سایر روش‌ها می‌باشد. همچنین (شکل ۲) نشان می‌دهد که در هر دو ایستگاه مقدار شاخص‌های ارزیابی در روش PCA به طور کاملاً مشخص دارای مقادیر با ارزش کمتر می‌باشد. لذا در آزمون‌های بعدی اثر سایر روش‌های انتخاب متغیر در ریزمقیاس نمایی مورد مقایسه قرار گرفت.

در (جدول ۵) مقایسه مشخصه‌های آماری بین روش‌های انتخاب متغیر انجام شده است. در هر کدام از مشخصه‌های آماری مقادیر با ارزش بالاتر با رنگ خاکستری مشخص شده اند. با استناد به مقادیر

تشخیص درست بودند. لکن مقایسه شاخص‌های دقت روش‌های مختلف در تشخیص سری‌های تر و خشک، نشان می‌دهد که روش ParCA دارای عملکرد بهتری نسبت به دو روش دیگر (CA و ParCA) می‌باشد.

مقابل تعداد روزهای تر و خشک ریزمقیاس شده تحت تأثیر روش‌های مختلف انتخاب متغیر، مشخص می‌شود که عملکرد CA در تشخیص درست روزهای تر مناسب نمی‌باشد و عملکرد دو روش SRA و ParCA بسیار مطلوب می‌باشد. به طوری که از ۷۸ روز تر موجود در آمار مشاهداتی، روش‌های SRA و ParCA به ترتیب دارای ۶۰ و ۶۴



شکل ۲- وضعیت شاخص‌های ارزیابی در ایستگاه بیرجند (a) و ایستگاه اردبیل (b)
 Figure 2- Efficiency criteria in Birjand (a) & Ardebil (b) stations

جدول ۵- مقادیر آماره‌های توصیفی در سری داده‌های مشاهداتی و ریزمقیاس شده

Table 5- Measured and downscaled descriptive statistics

آماره Statistic	ایستگاه Station	Observation	Downscaled		
			SRA	CA	ParCA
میانگین (میلی‌متر) Mean	اردبیل Ardebil	0.59	0.415	0.445	0.41
	بیرجند Birjand	0.254	0.624	0.068	0.599
میانه (میلی‌متر) Median	اردبیل Ardebil	0	0.43	0.473	0.48
	بیرجند Birjand	0.000	0.155	0.003	0.121
انحراف از معیار (میلی‌متر) Std. Div.	اردبیل Ardebil	1.766	0.148	0.094	0.163
	بیرجند Birjand	1.396	1.078	0.306	1.220
چولگی (میلی‌متر) Skewness	اردبیل Ardebil	3.998	-0.342	-1.709	-1.220
	بیرجند Birjand	7.807	2.547	11.898	3.289
کشیدگی (میلی‌متر) Kurtosis	اردبیل Ardebil	17.217	-0.197	3.291	0.193
	بیرجند Birjand	69.575	6.523	172.873	11.434

جدول ۶- مقایسه رخداد‌های تر و خشک در ایستگاه بیرجند

Table 6- comparison of dry and wet events in Birjand

ریزمقیاس شده Downscaled	مشاهداتی Observation	مجموع Summation		CSI	
		روز تر Wet day	روز خشک Dry day		
		روز تر Wet day	60		169
روز خشک Dry day	18	1214	1232		
CA	روز خشک Dry day	13	5	18	0.15
	روز خشک Dry day	65	1378	1443	
SRA	روز خشک Dry day	64	214	278	0.22
	روز خشک Dry day	14	1169	1183	
مجموع Summation		78	1383	-	

روش‌های مختلف در تشخیص روزهای تر در ایستگاه بیرجند بسیار بهتر از اردبیل می‌باشد. این امر با مقایسه شاخص عملکرد CSI مشخص می‌شود. مطالعه هارفام و ویلیبی (۱۲) که از شاخص‌های CSI جهت ارزیابی دقت روش‌های مختلف ریزمقیاس نمایی در

از ۲۰۰ روز تر ثبت شده در آمار مشاهداتی ایستگاه سینوپتیک اردبیل، روش SRA ۱۱ روز، روش CA ۱۸ روز و روش ParCA ۹ روز را تر تشخیص داده‌اند. در مقایسه با عملکرد روش‌های مختلف و اثر آن‌ها در ریزمقیاس نمایی، می‌توان بیان نمود که عملکرد

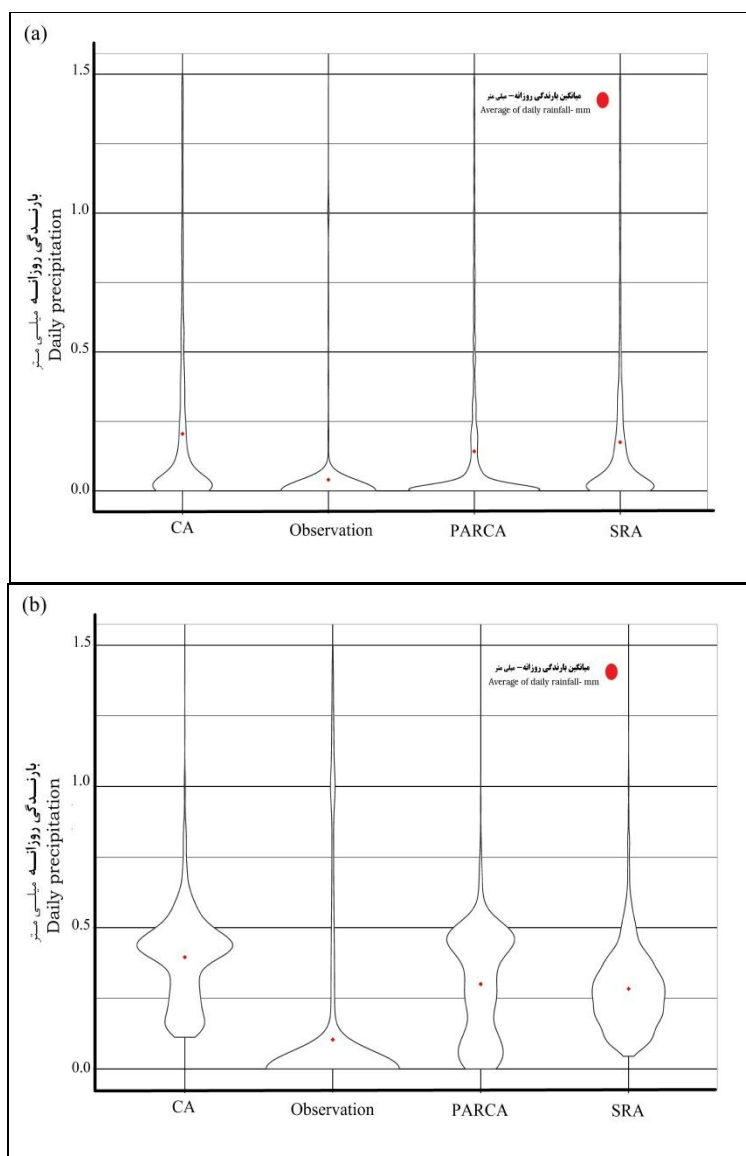
روش‌های انتخاب متغیر ریزمقیاس نمایی بارش در ایستگاه بیرجند مطلوبیت بیشتری نسبت به اردبیل داشت.

نتیجه‌گیری

مطالعات زیادی در ریزمقیاس نمایی آماری و دینامیکی انجام شده است که هدف آن‌ها ارتقای دقت خروجی مدل‌های GCMs می‌باشند تا از این طریق تخمین‌های مربوط به تغییرات مؤلفه‌های اقلیمی تخت تأثیر پدیده تغییر اقلیم بهبود پیدا نماید.

شبیه‌سازی بارش‌های سنگین کشور انگلستان انجام شده است، نیز با نتایج مطالعه هم‌خوانی دارد (۱۲).

با توجه به مقدار ضریب تغییرات بارش روزانه در دو ایستگاه بیرجند و اردبیل (به ترتیب ۱۸ و ۳۳ درصد) می‌توان بیان نمود که به دلیل اینکه تغییرات مقادیر بارش روزانه در مناطق نیمه‌خشک و سرد بسیار بیشتر از مناطق خشک می‌باشد، انتظار می‌رود که خطای دینامیک داده‌های این مناطق بیشتر از مناطق خشک می‌باشد. به طور مستقیم بالا بودن میزان خطای دینامیک داده‌ها موجب افزایش خطای سیستماتیک (مدل‌سازی یا در این مطالعه ریزمقیاس نمایی) خواهد شد (۲۲). از این رو در آزمون‌های به کار رفته در این مطالعه عملکرد



شکل ۳- نمودار Violin بارندگی مشاهداتی و ریزمقیاس شده در ایستگاه بیرجند (a) و ایستگاه اردبیل (b)
Table 3- Measured and downscaled violin plot in Birjand (a) and Ardebil (b)

جدول ۷- مقایسه احتمالاتی شاخص‌های رخداد در ایستگاه اردبیل

Table 7- comparison of dry and wet events in Ardebil

ریزمقیاس شده Downscaled	مشاهداتی Observation		مجموع Summation	CSI	
	روز تر Wet day	روز خشک Dry day			
	روز تر Wet day	9			191
روز خشک Dry day	37	1244	1261		
CA	روز خشک Dry day	18	182	200	0.07
	روز خشک Dry day	57	1204	1261	
SRA	روز خشک Dry day	11	189	200	0.05
	روز خشک Dry day	19	1242	1261	
مجموع Summation	200	1261	-		

به طوری که مقایسه مقدار شاخص‌های ارزیابی و مقایسه مشخصه‌های آماری بارندگی مشاهداتی و ریزمقیاس شده نشان داد که در ایستگاه بیرجند روش CA بهتر از سایر روش‌ها می‌باشد. اما مقایسه سری‌های تر و خشک نشان داد که دقت روش SRA بهتر از سایر روش‌ها می‌باشد. بررسی نتایج در ایستگاه اردبیل نشان داد که به دلیل دینامیک بیشتر داده‌های بارش روزانه نسبت به ایستگاه بیرجند، عملکرد روش‌های انتخاب متغیر در ایستگاه اردبیل از مطلوبیت کمتری برخوردار می‌باشد. نتایج مقایسه شکل توزیع آماری (پراکنش داده‌ها) و مقایسه سری‌های خشک و تر مشاهداتی و ریزمقیاس شده حاکی از عملکرد نامطلوب روش‌های انتخاب متغیر در ریزمقیاس‌های نمایی بارش روزانه اردبیل بود. در آزمون‌های مقایسه شاخص‌های ارزیابی و مشخصه‌های آماری نشان داد که روش‌های CA و SRA ضمن تشابه به یکدیگر، عملکرد بهتری نسبت به روش ParCA دارد.

در این پژوهش‌ها انتخاب متغیر ورودی به عنوان یکی از ابتدایی‌ترین گام ریزمقیاس‌نمایی، نقش بسیار مهمی در دقت شبیه‌سازی‌های اقلیمی دارد. در این مطالعه عملکرد روش‌های مختلف انتخاب متغیر در ریزمقیاس‌نمایی بارش روزانه در دو اقلیم متفاوت خشک- کویری (بیرجند) و سرد- نیمه‌خشک (اردبیل) مورد بررسی قرار گرفت. همچنین به منظور بررسی اثرات تغییر اقلیم از خروجی مدل CanESM2 تحت ریزمقیاس‌نمایی ANN استفاده شد.

پس از ارزیابی نتایج آزمون‌های متعدد مشخص شد که ریزمقیاس‌نمایی بارش روزانه در اقلیم‌های خشک- کویری عملکرد بهتری نسبت به اقلیم‌های سرد- نیمه‌خشک می‌باشد. به طوری که در تمامی آزمون‌های به کار رفته شده، دقت روش‌های انتخاب متغیر در بیرجند دارای مطلوبیت بیشتری می‌باشد. همچنین آزمون‌های مختلف نتایج متفاوتی را در خصوص عملکرد روش‌های انتخاب متغیر ارایه داد.

منابع

- Bengio Y. and LeCun Y., 2007. Scaling learning algorithms towards AI1-41. In Bottou L., Chapelle O., DeCoste D., Weston J. Large-scale kernel machines. MIT Press.
- Camici S., Palazzi E., Pier, A., Brocca L., Moramarco T., and Provenzale A. 2015. Comparison between dynamical and stochastic downscaling methods in central Italy. In EGU General Assembly Conference Abstracts, 10270.. 10270.
- Chanda K., and Maity R. 2018. Global Climate Pattern Behind Hydrological Extremes in Central India. In Climate Change Impacts (pp. 71-89). Springer, Singapore.
- Chen H., CY X., and Guo S.L. 2012. Comparison and evaluation of multiple GCMs, statistical downscaling and hydrological models in the study of climate change impacts on runoff. Journal of hydrology, 434, 36-45.
- Chen J., Brissette F.P., and Leconte R. 2011. Uncertainty of downscaling method in quantifying the impact of climate change on hydrology. Journal of Hydrology. 401:190-202.
- Devak M., Dhanya C.T., and Gosain A.K. 2015. Dynamic coupling of support vector machine and K-nearest neighbour for downscaling daily rainfall. Journal of Hydrology, 525:286-301.

7. Draper N.R., Smith H., and Pownell E. 1966. Applied regression analysis. 3: 217-220. New York: Wiley.
8. Fisher R.A. 1958. Statistical Methods for Research Workers, 13th Ed., Hafner.
9. Fistikoglu O., and Okkan U. 2011. Statistical Downscaling of Monthly Precipitation Using NCEP / NCAR Reanalysis Data for Tahtali River Basin in Turkey, Journal of Hydrologic Engineering, 16, 157-164.
10. Gaitan C. F., Hsieh W. W., and Cannon A. J. 2014. Comparison of statistically downscaled precipitation in terms of future climate indices and daily variability for southern Ontario and Quebec, Canada. Climate Dynamics, 43(12):3201-3217.
11. Hammami D., Lee T.S., Ouarda T.B. M.J., and Le J. 2012. Predictor selection for downscaling GCMs data with LASSO. Journal of Geophysical Research Atmospheres, 117(17):1-11.
12. Harpham C., and Wilby R.L. 2005. Multi-site downscaling of heavy daily precipitation occurrence and amounts. Journal of Hydrology, 312(1):235-255.
13. Hassan Z., Shamsudin S., and Harun S. 2014. Application of SDSM and LARS-WG for simulating and downscaling of rainfall and temperature. Theoretical and Applied Climatology. 116:243-257.
14. Hessami M., Gachon P., Ouarda T., and St-Hilaire A. 2008. Automated regression-based statistical downscaling tool. Environmental Modelling & Software, 23:813-834.
15. Huth R. 1999. Statistical downscaling in central Europe: evaluation of methods and potential predictors. Climate Research, 13:91-1011. IPCC (2001) Climate change 2001: impacts, adaptation and vulnerability. In: McCarthy JJ., Canziani oF, Leary NA., Dokken DJ., White KS (eds) Contribution of working group II to the third assessment report of the intergovernmental panel on climate change. Cambridge University Press, Cambridge
16. IPCC (2001) Climate change 2001: impacts, adaptation and vulnerability. In: McCarthy JJ., Canziani oF, Leary NA., Dokken DJ., White KS (eds) Contribution of working group II to the third assessment report of the intergovernmental panel on climate change. Cambridge University Press, Cambridge
17. IPCC (2007) Summary for policymakers. Climate change 2007: the physical science basis. In: Solomon S, Qin D, Manning M, Chen Z, Marquis M, Averyt KB, Tignor M, Miller HL (eds) Contribution of working group I to the fourth assessment report of the intergovernmental panel on climate change. Cambridge University Press, Cambridge, pp 1-18
18. Jafarzadeh A., Pourreza-Bilondi M., Khashei-Siuki A., Aghakhani A., and Yaghoobzadeh M. 2017. Reliability estimation of rainwater catchment system using future GCMs output data (case study: Birjand City). In 10th World Congress On Water Resources And Environment (pp. 536-541).
19. Kaviani M., and MirRokni S.M. 2014. Application of principal component analysis to meteorological data in ANN input selection. Geophysics Journal of IRAN, 9(1):71-84.
20. Khan M.S., Coulibaly P., and Dibike Y. 2006. Uncertainty analysis of statistical downscaling methods. Journal of Hydrology, 319(1-4):357-382. <https://doi.org/10.1016/j.jhydrol.2005.06.035>.
21. Khashei-Siuki A., Shahidi A., Pourreza-Bilondi M., Amirabdizadeh M., and Jafarzadeh A. 2018. Performance Assessment of ANN and SVM for downscaling of daily rainfall in dry regions. Iranian journal of Soil and Water Research. (in Persian)
22. Krause P., Boyle D.P., and Båse F. 2005. Comparison of different efficiency criteria for hydrological model assessment. Advances in Geosciences, 5:89-97.
23. McCulloch W.S., and Pitts W. 1943. A logical calculus of the ideas immanent in nervous activity. The Bulletin of Mathematical Biophysics, 5(4):115-133.
24. Meenu R., Rehana S., and Mujumdar P.P. 2013. Assessment of hydrologic impacts of climate change in Tungabhadra river basin, India with HEC-HMS and SDSM. Hydrological Processes, 27(11):1572-1589.
25. Najafi M.R., Moradkhani H., and Wherry S.A. 2011. Statistical Downscaling of Precipitation Using Machine Learning with Optimal Predictor Selection. Journal of Hydrologic Engineering, 16(8): 650-664.
26. Nasserli M., and Zahraie B. 2013. Performance Assessment of Different Data Mining Methods in Statistical Downscaling of Daily Precipitation. Journal of Hydrology, 492, 1-14.
27. Pervez M.S., Henebry G.M. 2014. Projections of the Ganges-Brahmaputra precipitation downscaled from GCMs predictors. Journal of Hydrology 517:120-134.
28. Raje D., Mujumdar P.P. 2011. A comparison of three methods for downscaling daily precipitation in the Punjab region. Hydrology Process, 25(23):3575-3589.
29. Rumelhart D.E., Hinton G.E., and Williams R.J. 1986. Learning representations by back-propagating errors. Nature, 323(6088):533-536.
30. Salathe E.P., Mote P.W., and Wiley M.W. 2007. Review of scenario selection and downscaling methods for the assessment of climate change impacts on hydrology in the United States Pacific Northwest. International Journal of Climatology, 27(12):1611-1621.
31. Semenov M. 2002. LARS-WG A Stochastic Weather Generator for Use in Climate Impact Studies. In *Version 3*, ed. R. Research.
32. Vousoughi F., Dinpashoh Y., and Aalami M. 2010. Effect of Drought on Groundwater Level in the Past Two

- Decades (Case study: Ardebil Plain). *Water and Soil Science*, 21(4):165-179. (in Persian)
33. Wilby R.L., and Wigley T.M.L. 2000. Precipitation predictors for downscaling: observed and general circulation model relationships. *International Journal of Climatology*, 20(6):641-661.
 34. Wilby R.L., Dawson C.W., and Barrow E.M. 2002. SDSM—a decision support tool for the assessment of regional climate change impacts. *Environmental Modelling & Software*, 17(2):145-157.
 35. Wilks D.S., 1995. *Statistical Methods in the Atmospheric Sciences*. Academic Press, San Diego, California
 36. Wood A.W., Leung L.R., Sridhar V., and Lettenmaier D.P. 2004. Hydrologic implications of dynamical and statistical approaches to downscaling climate model outputs. *Climatic Change* 62: 189–216.
 37. Yang C. 2016. Performance comparison of three predictor selection methods for statistical downscaling of daily precipitation. *Theoretical and Applied Climatology*. 131(1-2):43-54.



Examination of Feature Selection Methods for Downscaling of Daily Precipitation in Two Different Climates

J.RamezaniMoghadam^{1*} - M.Yaghoubzadeh²- A. Jafarzadeh³

Received: 22-05-2018

Accepted: 29-07-2018

Introduction & Background: Assessment of climate change impacts on hydrology is relied on the information of climate changes in adequate scale. Due to outputs of GCMs (General Circulation Models) that are the most confident tools for simulating climate change impacts but are available in coarse resolution. Downscaling process which is classified to several methods such as transfer function, weather generator and weather typing is performed for improving of GCMs projection and using them in local scale. Meanwhile feature selection is the main essential step in downscaling with transfer function. Because the main goal of downscaling is the improvement of GCMs projections, several researches examined vary approaches for feature selection. This study aims to assess performance of downscaling daily precipitation under four different selection methods such as PCA, CA, SRA and ParCA using comprehensive comparison tests.

Materials and Methods: Measured daily rainfall for Ardebil (with cold semi-arid climate) and Birjand (arid climates) were collected for the period from 1977 to 2004. The CanESM2 (Canadian Earth System Model) outputs were used as GCM for simulating of climate change impacts on precipitation pattern. So of CanESM2 outputs (large scale predictors) and measured daily precipitation (local scale predictants) were considered as input and target for downscaling respectively. The Artificial Neural Network (ANN) which widely has been used in climate change researches was selected as downscaling method. Despite of the most of literature have used only efficiency criteria for distinguishing from different approaches in downscaling, this study reveals performance of feature selection methods based on either them or statistical tests. The comparison tests between measured and downscaled rainfall such as assessment criteria, statistics characteristics comparison, contingency table event for wet and dry series diagnostics and Violin plot were used as tools for skill assessment of feature selection approaches.

Results and Discussion: Results showed that although different methods of predictor selection had includes various subsets, predictors such as relative humidity at surface and zonal velocity component at 500-hPa pressure levels in Birjand and mean temperature at 2m, mean sea level pressure and rotation of the air in Ardebil are the most descriptive features which have more relationship with measured daily precipitation. The efficiency criteria of comparing measured and downscaled precipitation indicated that CA method is superior to other in Birjand station and SRA's results were better than those of other in Ardebil station. Value of RMSE, R and NSE was achieved 1.2 mm/day, 0.55 and 0.25 in Birjand and 1.75 mm/day, 0.14 and 0.013 in Ardebil respectively. The examination of measured and downscaled statistical characteristics reveals that CA has the better influence on downscaling than those of others in Birjand station. In this comparative test most of downscaled statistical components such as mean, median and skewness under CA have more similarity to measured values. But in Ardebil, with cold and arid climate, performance of SRA to downscale was the same as performance of CA to it. Also both SRA and CA were better than ParCA. The skill assessment of different methods to fit measured and downscaled variability by violin plot showed that generally ParCA outperformed other method in Birjand station. The comparison of violin plots, in Ardebil, revealed that no one of predictor selection methods has acceptable accuracy for fitting measured variability. Outcomes of contingency table event showed although all feature selection methods have not remarkable capability for distinguishing from the measured wet and dry series in Ardebil station, performance of ParCA and SRA were acceptable in Birjand station. The values of CSI for ParCA and SRA were calculated 0.25 and 0.22 in Birjand and it shows that more of 20 percent of ParCA and

1- Assistant Professor of Water Engineering Department, Faculty of Agriculture and Natural Resources, University of Mohaghegh Ardabili

(*-Corresponding Author Email:j_ramezani@uma.ac.ir)

2 and 3- Assistant Professor and PhD Student of Water Resource Engineering, Department of Science and Water Engineering, University of Birjand

SRA's diagnostics was correct.

Conclusions: By assessing of results, it can be inferred that generally downscaling of daily rainfall in Birjand station is outperforming Ardebil. In other expression daily downscaling of precipitation in arid climate has better results than cold and arid climate. Also different tests have various results about feature selection methods. In Ardebil, SRA in efficiency criteria test and both SRA and CA in statistics characteristics have better performance than others. But in this region no methods have remarkable performance in violin and dry and wet tests.

Keywords: CanESM2, Partial Correlation, PCA, Reduction Dimensions, Wet and Dry Spell