

مقایسه عملکرد مدل های BCSD، ANN، GP و SVM در شبیه سازی دما

سید مصطفی طباطبائی^{۱*}، محمد ناظری تهرودی^۱، مهدی دستورانی^۲

^۱ دانشجوی دکتری منابع آب دانشکده کشاورزی دانشگاه بیرجند، بیرجند، ایران
^۲ استادیار گروه علوم و مهندسی آب دانشکده کشاورزی دانشگاه بیرجند، بیرجند، ایران

تاریخ دریافت: ۱۳۹۶/۱۲/۱۸، تاریخ پذیرش: ۱۳۹۷/۰۳/۲۱

چکیده

در این پژوهش به منظور ارزیابی عملکرد مدل های شبیه سازی از چهار روش BCSD، ANN، GP و SVM استفاده شد. مدل سازی بر اساس داده های بزرگ مقیاس مدل های گردش عمومی جو و دمای میانگین روزانه ایستگاه سینوپتیک اهواز در دوره ۲۰۰۴-۱۹۷۱ انجام شد و ارزیابی هر مدل بر اساس معیارهای ضریب همبستگی و خطای مدل سازی بین داده های مشاهداتی و شبیه سازی شده انجام گرفت. نتایج مدل سازی نشان داد که ضریب همبستگی بین داده های مشاهداتی و شبیه سازی شده در روش SVM از سایر روش ها بیشتر بوده و مقدار آن ۰/۹۹۶۰ می باشد. ضریب همبستگی برای روش ANN، GP و BCSD به ترتیب برابر ۰/۹۳۹۳، ۰/۹۳۸۳ و ۰/۴۹۳۶ می باشد. همچنین نتایج حاصل از ارزیابی خطای شبیه سازی با استفاده از معیارهای RMSE و NSE برای SVM به ترتیب ۰/۶۷۷ و ۰/۹۹۵ درجه سانتی گراد محاسبه شده است. به طور مشابه این مقادیر برای GP ۱/۶۴۴ و ۰/۹۶۹، ANN ۱/۶۵۷ و ۰/۹۶۸ و BCSD ۶/۱۷۴ و ۰/۶۶۱ درجه می باشد. بنابراین SVM در مدل سازی دمای میانگین نسبت به سایر روش ها از عملکرد بهتری برخوردار است و مدل سازی دما به روش BCSD نسبت به سایر روش ها از دقت کمتری برخوردار است. روش GP نسبت به ANN برتری ضعیفی دارد و پیشنهاد می شود برای ارزیابی عملکرد دقیق تر این دو مدل از دماهای حداقل و حداکثر استفاده شود.

کلمات کلیدی: برنامه ریزی ژنتیک، تغییر اقلیم، مدل سازی، ماشین بردار پشتیبان

اقليم، آميخته‌اي از ويژگي‌هاي چيره شده و ماندگار جويي يك گستره‌ي جغرافيايي در گذر زمان است و اغلب بر اساس متغيرهايي مانند دما، بارش، رطوبت، وزش باد، تابش خورشيد، تعداد روزهاي آفتابي، دماي سطح دريا و ضخامت لايه‌هاي يخ در آب دريا تعيين مي‌شود. مجموعه‌ي اين عوامل در بلند مدت همراه با ديگر ويژگي‌هاي منطقه‌اي مانند مدت زمان فصل كشت و شدت سيلاب‌ها، اقليم يك منطقه را تعيين مي‌کنند (Salinger et al, 2000). در دهه‌هاي اخير، افزايش مصرف سوخت‌هاي فسيلي به منظور تأمين انرژي بخش‌هاي اقتصادي، از يك سو و جنگل زدائي و تخریب محیط زیست از سويي ديگر باعث افزايش روز افزون گازهاي گلخانه‌اي در سطح کره زمين شده است. پس از بخار آب، CO₂ مهم‌ترين گاز گلخانه‌اي جذب کننده‌ي اشعه مادون قرمز در اتمسفر است (علي پور، ۱۳۹۳). افزايش غلظت CO₂ موجب افزايش درجه‌ي حرارت اتمسفر کره‌ي زمين مي‌شود. به دنبال گرمایش جهانی، گردش عمومي جو نيز تحت تأثير قرار مي‌گيرد و سبب تغييراتي در مقدار، شدت، مدت و زمان وقوع بارش در نقاط مختلف کره زمين مي‌شود. لذا مجموعه‌ي اين عوامل پديده‌ي تغيير اقليم را شکل مي‌دهند (Vanuytrecht et al, 2012).

پديده تغيير اقليم و تأثيرات آن به عنوان يکي از مهم‌ترين چالش‌هاي پيش رو در بخش‌هاي کشاورزي و منابع آب است. هيات بين المللي تغيير اقليم در سال ۲۰۰۱ گزارش داد که اقليم در حال تغيير و گرمایش جهانی در حال وقوع است به طوري که ميانگين دما در جهان از سال ۱۸۶۱ ميلادي روندی افزايشی داشته و حتی در قرن بيستم ميلادي نيز اين افزايش بين (۰/۶ ± ۰/۲) درجه سانتی گراد بوده است (IPCC, 2001). طبق گزارش IPCC بر اساس تخمين تجمع گازهاي گلخانه‌اي، افزايش دمائي معادل ۳/۵ درجه سانتی گراد تا سال ۲۱۰۰ ميلادي پيش بيني مي‌شود و مقدار گرم شدن

زمين طی صد سال آینده، بيش از آن چه طی ۱۰۰۰۰ سال گذشته رخ داده است، خواهد بود (نادري و همکاران، ۱۳۹۶). افزايش دما بر ميزان بارندگي و رطوبت خاک و شدت خشک سالي‌ها تأثير گذار بوده و مي‌تواند عملکرد و توليد محصولات کشاورزي را تحت تأثير قرار دهد (Feng and Liu, 2015).

با توجه به اهميت پيش بيني‌هاي هر چه دقيق‌تر تغيير در شرايط جوي و نقش مدل‌هاي مورد استفاده در شبیه‌سازي و پيش بيني پارامترهاي موثر بر چرخه هيدرولوژي پژوهش‌هاي مختلفی پيرامون مقايسه عملکرد مدل‌هاي مختلف انجام شده است. (Adamowski, 2013) براي شبیه‌سازي بارندگي در يك حوضه کوهستاني مدل‌هاي SVM و ANN را مورد ارزيابي قرار داد. نتايج وي نشان دهنده برتري SVM بود. (Tezel and Buyukyildiz 2015) کارآيي ANN و SVM را در برآورد تبخیر ماهانه مورد ارزيابي قرار دادند. در پژوهش آن‌ها دما، رطوبت نسبي، سرعت باد، و بارش به عنوان ورودی و تبخیر به عنوان خروجی مدل در نظر گرفته شد. نتايج آن‌ها نشان داد که هر دو مدل عملکرد مشابهی دارند. (Ghorbani et al 2016) مدل SVM و ANN را در پيش بيني جريان رودخانه مورد بررسي قرار دادند. نتايج آن‌ها نشان دهنده عدم قطعيت کمتر در خروجي SVM بود. (Sedighi et al 2016) جهت پيش بيني فرآيند بارش- رواناب با استفاده از اطلاعات سنجنده موديس طی سال آماری ۲۰۰۳-۲۰۰۵ از SVM و ANN بهره بردند. طبق نتايج آن‌ها SVM نسبت به ANN نتايج بهتري ارائه نموده است. (Kisi 2016) به منظور مدل‌سازي تبخیر ميزان دقت سه روش SVM، رگرسيون چند متغيره تطبيقي و مدل درختي را بررسي کرد. در مرحل اول ميزان تبخیر با استفاده از داده‌هاي محلي برآورد شد که نتايج حاکی از عملکرد بهتر مدل SVM نسبت به دو مدل ديگر بود. در مرحل دوم ميزان تبخیر بدون داده‌هاي محلي برآورد شد که نتايج مدل رگرسيون چندمتغيره تطبيقي، نسبت به دو مدل ديگر

روزانه ثبت شده در ایستگاه سینوپتیک شهر اهواز در دوره (۲۰۰۴-۱۹۷۱) و ارائه مدل برتر می‌باشد. برای دستیابی به هدف مورد نظر ضریب همبستگی بین داده‌های مشاهداتی و شبیه‌سازی شده به همراه خطای مدل سازی مورد بررسی قرار گرفته است.

مواد و روش‌ها

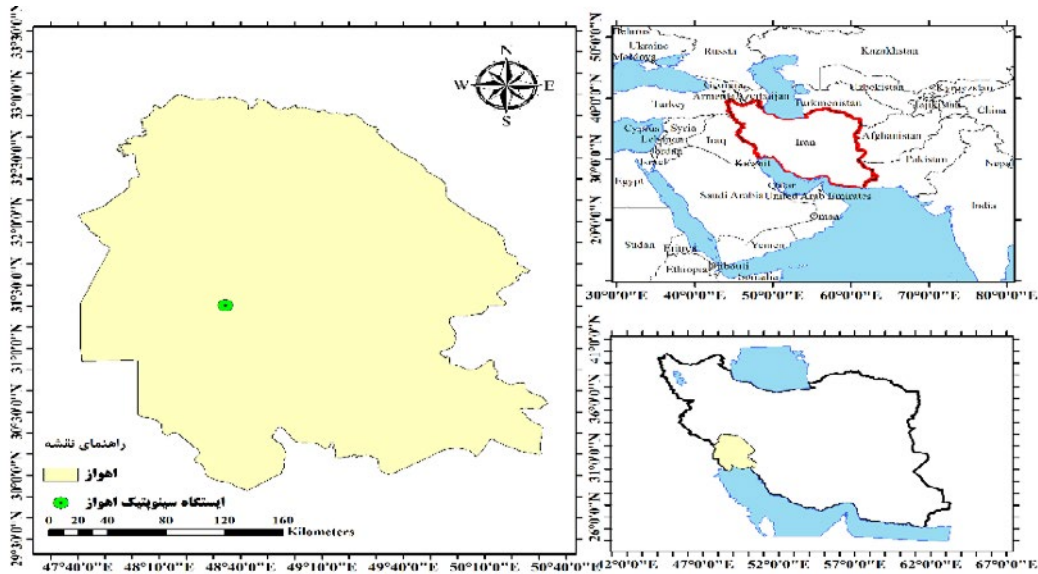
معرفی منطقه

استان خوزستان به عنوان قطب کشاورزی ایران شناخته شده است و شهر اهواز به عنوان مرکز این استان دارای مساحتی معادل ۱۸۶۵۰ هکتار می‌باشد. جمعیت این شهر طبق آمار سرشماری سال ۱۳۹۵ برابر ۱۳۰۲۵۹۱ نفر می‌باشد و در طول جغرافیایی ۴۰ درجه و ۴۸ دقیقه و عرض ۳۱ درجه و ۲۰ دقیقه در بخش جلگه ای استان خوزستان با ارتفاع ۲۲/۵ متر از سطح دریا واقع شده است. رودخانه کارون به عنوان پرآب‌ترین رودخانه ایران از مرکز این شهر عبور می‌کند و شهر اهواز را به دو بخش شرقی و غربی تقسیم می‌کند. موقعیت ایستگاه سینوپتیک اهواز در شکل (۱) نشان داده شده است.

در این پژوهش از مقادیر مشاهداتی متوسط دمای روزانه در دوره زمانی ۱۹۷۱-۲۰۰۴ و خروجی پارامترهای شبیه‌سازی شده NCEP برای مدل سازی استفاده شد. داده‌های NCEP مشتمل بر ۲۶ متغیر و شامل پارامترهایی مانند میانگین دما در ارتفاع دو متری، میانگین فشار سطح دریا، رطوبت نسبی و رطوبت ویژه در سطح‌های ۵۰۰ و ۸۵۰ هکتوپاسکال و نزدیک سطح زمین و غیره می‌باشد. پس از مرتب‌سازی داده‌ها متغیر دما را به عنوان متغیر وابسته و داده‌های مدل گردش عمومی NCEP به عنوان متغیر مستقل به مدل معرفی شد. دوره آماری ۱۹۹۶-۱۹۷۱ به عنوان دوره آموزش و دوره ۲۰۰۴-۱۹۹۷ به عنوان دوره آزمون یا تست در نظر گرفته شدند. با در نظر گرفتن مقادیر مشاهده شده و شبیه‌سازی شده توسط هر یک از مدل‌ها بعد از بررسی ضریب همبستگی، ارزیابی بر اساس دو معیار RMSE و NSE انجام شد.

دقت بیشتری از خود نشان داد. دهقانی و همکاران (۱۳۹۶) به مقایسه عملکرد مدل‌های GP، SVM و شبکه بیزین برای پیش‌بینی جریان رودخانه پرداختند. نتایج آن‌ها نشان داد که مدل SVM با ضریب همبستگی ۰/۸۱، ریشه میانگین مربعات خطای ۰/۰۰۲ متر مکعب نسبت به GP و شبکه بیزین از عملکرد بهتری برخوردار بوده است. سلگی (۱۳۹۷) برای پیش‌بینی بارش ماهانه با استفاده از داده‌های دما و رطوبت نسبی ماهانه در یک دوره ۳۲ ساله (۱۳۹۳-۱۳۶۲) از دو مدل GP، SVM بهره برد. نتایج وی نشان داد که ضریب همبستگی هر دو مدل ۰/۹۲ و RMSE این دو مدل به ترتیب ۰/۴۸۶ و ۰/۴۷۸ بوده است. کدخدا حسینی و همکاران (۱۳۹۷) عملکرد دو مدل GP و SVM را در بازسازی داده‌های گمشده بارندگی در دوره (۱۳۸۹-۱۳۷۰) مورد بررسی قرار دادند. نتایج آن‌ها نشان داد که ضریب همبستگی در مدل GP و SVM به ترتیب ۰/۹۳ و ۰/۹۵ بود. ارزیابی خطا بر اساس معیار RMSE برای دو مدل به ترتیب ۱۳ و ۱۲/۲ میلی‌متر محاسبه شده است. شاهی نژاد و دهقانی (۱۳۹۷) به مقایسه مدل سازی با استفاده از SVM و ANN در تخمین اکسیژن محلول آب رودخانه در یک دوره ۱۰ ساله (۲۰۱۶-۲۰۱۰) پرداختند. نتایج آن‌ها نشان داد که ANN با ضریب همبستگی ۰/۹۶ و RMSE برابر ۰/۶۶۸ دارای بیشترین ضریب همبستگی و کمترین خطا بوده است.

با توجه به تاثیر پدیده تغییر اقلیم بر پارامتر دما و اثرگذاری تغییرات دما بر چرخه هیدرولوژیکی، پیش‌بینی و مدل‌سازی هر چه دقیق‌تر دما برای دوره‌های آتی جهت ارزیابی تاثیر تغییرات دمایی بر بارندگی و منابع آب سطحی و زیرزمینی گامی موثر جهت موفقیت در اجرای برنامه‌ریزی‌های آینده محسوب می‌شود و نوع مدل‌های مورد استفاده بر میزان دقت مدل‌سازی تاثیرگذار است. هدف از انجام این پژوهش مقایسه عملکرد مدل‌سازی دمای هوا به چهار روش GP، ANN، BCSD و SVM با استفاده از داده‌های بزرگ مقیاس مدل‌های گردش عمومی جو و سری داده‌های دمای



شکل ۱: موقعیت ایستگاه سینوتیک اهواز

ارتباط دهنده بین متغیرهای ورودی و خروجی بوده و لذا قادر به انتخاب خودکار متغیرهای مناسب مدل و حذف متغیرهای غیر مرتبط است که این امر سبب کاهش ابعاد متغیرهای ورودی خواهد شد. انتخاب ورودی‌های مناسب مدل یکی از مهم‌ترین مواردی است که بایستی در این روش مورد توجه قرار گیرد. این امر در شرایطی که از داده‌های ورودی ثانویه نیز بهره برده شود، از اهمیت مضاعفی برخوردار خواهد بود، زیرا ارائه داده‌های ورودی غیر مرتبط، سبب کاهش دقت مدل و ایجاد مدل‌های پیچیده‌تری می‌شود (Chen, 2003). فرآیند گام به گام برنامه ریزی ژنتیک به صورت مراحل زیر می‌باشد.

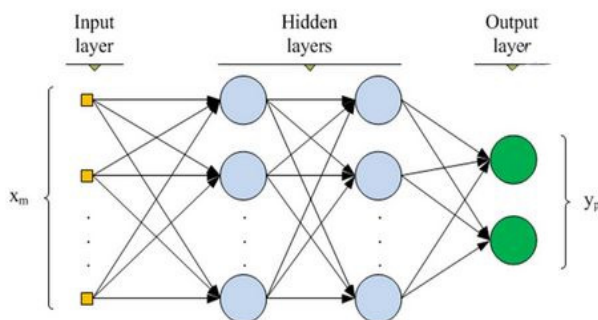
در مرحله اول یک جمعیت اولیه از توابع مرکب نشان دهنده مدل‌های پیش‌بینی، به صورت تصادفی در نظر گرفته می‌شود. سپس هر یک از افراد جمعیت مذکور با استفاده از توابع برازش، مورد ارزیابی قرار می‌گیرند. در مرحله سوم در هر تولید، مراحل زیر برای انتخاب یک جمعیت جدید دنبال می‌شود. الف: یکی از عملگرهای عبور، جهش و کپی انتخاب می‌شود. ب: تعداد مناسبی از افراد جمعیت حاضر انتخاب می‌شوند. ج: از عملگر انتخابی برای تولید فرزند استفاده می‌شود. د: فرزند یاد

معرفی مدل‌ها

برنامه ریزی ژنتیک^۱ (GP)

برنامه ریزی ژنتیک اولین بار توسط کوزا توسعه داده شد. این روش جزء روش‌های تکاملی محسوب می‌شود که بر مبنای نظریه داروین استوار است (Koza, 1992). برنامه ریزی ژنتیک بر خلاف الگوریتم ژنتیک روی ساختار درختی فرمول‌ها به جای سلسله ارقام دودویی عمل می‌کند و ساختارهای درختی از مجموعه توابع (عملگرهای ریاضی مورد استفاده در فرمول‌ها) و ترمینال‌ها (متغیرهای مسئله و اعداد ثابت) ایجاد می‌شوند. الگوریتم‌های یاد شده اقدام به تعریف یک تابع هدف در قالب معیارهای کیفی نموده و سپس تابع یاد شده را برای مقایسه جواب‌های مختلف حل مسئله در یک فرآیند گام به گام تصحیح ساختار داده‌ها به کار می‌گیرند و در نهایت، جواب مناسب را ارائه می‌نمایند (Ferreira, 2001).

در این روش، ابتدا بلوک‌های موجود که شامل متغیرهای ورودی و هدف و نیز تابع ارتباط دهنده آن‌ها می‌باشد، تعریف شده و سپس ساختار مناسب مدل و ضرایب آن تعیین می‌شوند. این روش شامل یک معادله



شکل ۲: ساختار یک شبکه عصبی مصنوعی

شبکه عهده‌دار است. چگونگی اتصال سلول‌های عصبی در لایه‌های مختلف، مشخص‌کننده ساختار شبکه است که معماری شبکه نام دارد. شبکه عصبی از چندین لایه (لایه‌های ورودی، لایه‌های مخفی و لایه‌های خروجی) تشکیل شده است. لایه‌ها وظیفه دریافت داده‌ها، پردازش و تولید کمیت خروجی را به عهده دارند. شکل (۲) نشان دهنده ساختار یک شبکه عصبی مصنوعی، با چند لایه ورودی، مخفی و خروجی می‌باشد. (Thierry et al, 2008)

مراحل طراحی یک مدل شبکه عصبی به منظور پیش‌بینی یا تخمین شامل: شناخت متغیرهای ورودی و خروجی، تبدیل مقادیر ورودی و خروجی به دامنه صفر تا یک، انتخاب هندسه مناسب برای شبکه عصبی، آموزش با داده‌های معرف، آزمون شبکه با داده‌های مستقل از مجموعه آموزشی و در صورت لزوم ادامه آموزش شبکه و پارامترهای آن می‌باشد. (Anagu et al, 2009)

انجام شبیه‌سازی با استفاده از شبکه عصبی در نرفازار برنامه‌نویسی متلب پس از معرفی داده‌های مشاهداتی و آموزش بخشی از داده‌ها، مرحله صحت‌سنجی و آزمایش توسط بخش دیگری از داده‌ها انجام شد و خروجی به صورت داده‌های شبیه‌سازی شده حاصل گردید.

جداسازی فضایی تصحیح‌اریبی^۳ (BCSD)

روش BCSD یکی از روش‌های ریزمقیاس‌نمایی می‌باشد که در ابتدا برای ریزمقیاس‌نمایی مکانی و زمانی پیش‌بینی مدل‌های هیدرولوژی به کار برده شد.

شده در یک جمعیت جدید وارد می‌شود. مدل مورد نظر با استفاده از برازش‌های مختلف مورد ارزیابی واقع می‌شود. و در مرحله چهارم تا رسیدن به حداکثر تعداد تولید، تکرار خواهد شد. (Borelli et al 2006). انجام این مراحل که به صورت کلی بیان شد در بسته نرم‌افزاری Genetic Programing ارائه شده است که مدل‌سازی انجام شده در این پژوهش توسط این نرم‌افزار انجام شده است.

شبکه عصبی مصنوعی^۲ (ANN)

شبکه‌های عصبی جزء آن دسته از سیستم‌های دینامیکی هستند که با پردازش روی داده‌های تجربی، دانش یا قانون نهفته در ورای داده‌ها را به ساختار شبکه منتقل می‌کنند و بر اساس محاسبات روی داده‌های عددی یا نمونه‌ها، قوانین کلی را فرا می‌گیرند. شبکه‌های عصبی مصنوعی می‌توانند با تحلیل اطلاعات، ارتباطات موجود بین آن‌ها را استخراج کرده و با به کار بستن آن در ازای یک سری از اطلاعات جدید مقادیر متناظر آن را تخمین بزنند. هر شبکه عصبی مصنوعی از عناصر پردازش که همان نرون‌های مصنوعی باشند تشکیل شده است که این نرون‌ها می‌توانند به روش‌های مختلف برای شکل دادن ساختار شبکه سازماندهی شوند (منهاج، ۱۳۹۳).

نرون به عنوان کوچکترین واحد پردازشگر داده‌ها در یک شبکه عصبی مصنوعی، اساس عملکرد و رفتار آن را تشکیل می‌دهد. از ترکیب چند نرون سلول ساخته می‌شود که بسته به نوع سلول وظیفه خاصی را در

رگرسیون بردار پشتیبان هدف پیدا کردن تابع $f(x)$ است به طوری که برای همه داده های آموزشی حداکثر به میزان ε از هدف های به دست آمده واقعی y_i انحراف دارد و در عین حال تا حد ممکن هموار است. تابع خطی f به شکل معادله (۱) بیان می شود که در آن عبارت (ω, x) بیانگر ضرب نقطه ای در R^d است (Basak et al, 2007).

$$F(x) = (\omega, x) + b, \omega \in R^d, b \in \mathbb{R} \quad (1)$$

هموار بودن در رابطه بالا به معنی ω کوچک است. بدین منظور لازم است تا $\|W^2\|$ حداقل شود. زمانی که برای مقدار واقعی f چندین خطا با دقت ε مجاز باشد از متغیرهای کمکی ξ_i و ξ_i^* برای غلبه بر قیود عملی نشدن مساله معرفی می شوند. و به صورت معادله (۲) تعریف می شود.

$$\text{Minimize } \frac{1}{2} \|\omega\|^2 + C \sum_{i=1}^l (\xi_i + \xi_i^*) \quad (2)$$

مقدار ثابت $C \geq 0$ توازنی بین هموار بودن f و خطای تجربی برقرار می کند (Vapnik, 1984). برای کاربرد ماشین های بردار پشتیبان در مسائل رگرسیونی مطابق معادله (۳) از تابع جدیدی به نام ε -insensitive استفاده کرد. این تابع خطاهایی که در یک فاصله معین از مقادیر واقعی هستند را نادیده می گیرد.

$$|\xi|_{\varepsilon} = \begin{cases} 0 & \text{if } |\xi| < \varepsilon \\ |\xi| - \varepsilon & \text{otherwise} \end{cases} \quad (3)$$

این تابع، خطای کمتر از \hat{a} را در نظر نمی گیرد. تابع خطا ε -insensitive به عنوان خطای تجربی جایگزین می شود. فرمول بندی دوگانه جهت توسعه بردار پشتیبان توابع غیرخطی استفاده می شود. با محاسبه ضرایب لاگرانژ تابع رگرسیون به صورت معادله (۴) و (۵) بهینه می شود.

$$\hat{u} = \sum_{i=1}^l (a_i - a_i^*) x_i \quad (4)$$

$$f(x) = \sum_{i=1}^l (a_i - a_i^*) x_i (x_i, x) + b \quad (5)$$

این روش دارای دو مرحله تصحیح اریبی و تولید داده می باشد در مرحله تصحیح اریبی، مقادیر میانگین گیری شده ماهانه مدل گردش عمومی (GCM) در دوره پایه با سری زمانی داده های مشاهده شده در مقیاس مکانی و دوره زمانی مشابه مقایسه می شوند. سپس شبیه سازی های GCM در دوره آینده بر اساس نتایج مقایسه، تنظیم و تصحیح می شوند. در انتها به روش تولید داده \hat{a} سری زمانی داده های روزانه اقلیم آینده به روش تصادفی تولید می شوند. در این روش پیش بینی شونده و پیش بینی کننده یکی هستند، با این تفاوت که مقیاس مکانی و زمانی آن ها مختلف است، به عنوان مثال از داده های شبکه ای دمای مدل گردش کلی برای تصحیح داده ها و سپس ریزمقیاس نمایی در مقیاس ایستگاه به روش تولید داده استفاده می شود. پس از انجام تصحیحات مرحله اول بر روی برون داد شبکه ای مدل گردش کلی، مرحله دوم شبیه سازی به روش BCSD انجام می شود (Ahmed et al, 2013).

ماشین بردار پشتیبان^۴ (SVM)

در شبکه های عصبی MLP و RBF غالباً توجه بر بهبود ساختار شبکه عصبی است، به نحوی که خطای تخمین شبکه عصبی کمینه شود. اما در نوع خاصی از شبکه عصبی، موسوم به ماشین بردار پشتیبان صرفاً بر روی کاهش ریسک عملیاتی مربوط به عدم عملکرد صحیح تمرکز می شود. ساختار یک شبکه SVM، اشتراکات زیادی با شبکه عصبی پرسپترون چند لایه^۵ (MLP) دارد و تفاوت اصلی آن در عمل در شیوه یادگیری است. ماشین بردار پشتیبان توسط (Vapnik and Chervonenkis, 1974; Vapnik, 1984) بر پایه تئوری یادگیری آماری توسعه داده شد و یکی از روش های یادگیری نظارت شده است که از آن برای طبقه بندی و رگرسیون استفاده می کنند. جزییات رگرسیون بردار پشتیبان به صورت زیر می باشد. در

4- Weather Generator
5- Support Vector Machines
6- Multi-Layer Perceptron

سیگموئید هستند از مزایای تابع کرنل می‌توان به حل مسائل محاسباتی که دارای ابعاد زیادی هستند امکان استفاده از ابعاد نامتناهی و کارآمد بودن از نظر زمانی و حافظه اشاره کرد. کرنل مورد استفاده در این پژوهش از نوع شعاعی پایه ای می‌باشد. با توجه به اینکه الگوریتم خطی SVM فقط به ضرب نقطه ای $\phi(x_i)$ و $\phi(x_j)$ بستگی دارد. با جایگزینی توابع کرنل به جای این ضرب‌های نقطه ای، تابع رگرسیون خطی در فضای جدید به صورت معادله (۶) نوشته می‌شود (اسکندری و همکاران، ۱۳۹۱).

$$f(x) = (w, \phi(x)) + b = \sum_{i=1}^n (a_i - a_i^*) (\phi(x_i), \phi(x)) + b$$

$$\rightarrow f(x) = \sum_{i=1}^n (a_i - a_i^*) k(x_i, x) + b \quad (6)$$

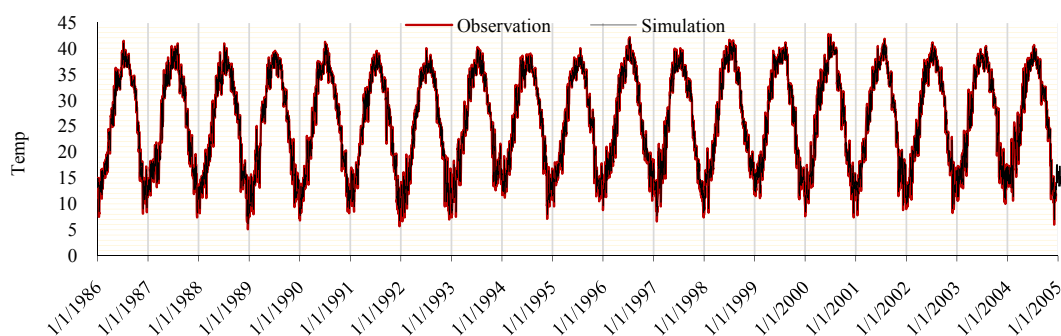
ارزیابی خطای مدل سازی

دو معیار ناش-ساتکلیف (NSE) ارائه شده از (Nash and Sutcliffe 1980) به عنوان نسبت خطای میانگین مربعات به انحراف معیار داده‌های مشاهداتی (معادله ۷) و ریشه میانگین مربعات خطا (RMSE) به عنوان بزرگی خطا (رابطه ۸) (Hyndman and Koehler 2006) برای ارزیابی میزان خطای هر مدل استفاده شد.

$$NSE = 1 - \frac{\sum_{i=1}^n (x_{obs} - x_{model})^2}{(x_{model} - \bar{x}_{model})^2} \quad (7)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (x_{obs} - x_{model})^2}{n}} \quad (8)$$

این معادله توسعه بردار ماشین نامیده می‌شود. یعنی w می‌تواند به طور کامل به صورت یک ترکیب خطی از نمونه‌های آموزشی توصیف شود. همچنین این رابطه نشان می‌دهد که نقطه‌هایی که ضریب لاگراژ غیر صفر دارند (یعنی نقطه‌هایی که روی خطوط و خارج از محدوده e-insensitive قرار می‌گیرند) در محاسبات $f(x)$ نقش دارند. این نقطه‌ها بردار پشتیبان نامیده می‌شوند. در روش SVM در حالتی که نتوان یک تابع رگرسیون خطی در فضای ورودی به داده‌ها برازش داد از یک نگاهت غیرخطی ϕ برای تبدیل داده‌ها به یک فضای بالاتر استفاده می‌شود و سپس در این فضای جدید الگوریتم SVM استاندارد اجرا می‌شود و رگرسیون خطی در این فضای جدید متناظر با رگرسیون غیرخطی در فضای ورودی خواهد بود. انجام محاسبات در فضای با ابعاد زیاد با محدودیت‌هایی مواجه است. برای غلبه بر این مشکل از تابع کرنل استفاده می‌شود. این تابع یک جداکننده خطی متکی به ضرب داخلی بردارها است که به صورت $K(x_i, x_j) = x_i^T x_j$ می‌باشد. اگر نقطه‌های x_i و x_j با استفاده از انتقال ϕ به $\phi(x)$ به فضای با ابعاد بزرگتر انتقال یابد ضرب داخلی آن‌ها به صورت $K(x_i, x_j) = \phi(x_i)^T \cdot \phi(x_j)$ تبدیل خواهد شد. بنابراین از محاسبه x خودداری کرده و به جای آن از تابع کرنل استفاده می‌شود. انتخاب کرنل‌های غیرخطی اجازه ساخت جداکننده‌های خطی در صورتی که در فضای اصلی غیرخطی هستند را می‌دهد. متداول‌ترین توابع کرنل توابع چندجمله‌ای، توابع پایه شعاعی گوسی، توابع پایه شعاع نمایی و توابع

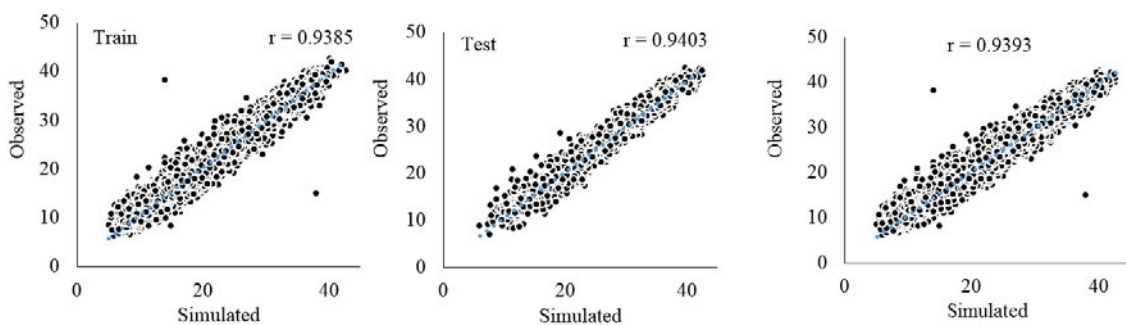


شکل ۳: مقایسه مقادیر مشاهداتی و شبیه‌سازی شده دمای متوسط به روش برنامه ریزی ژنتیک

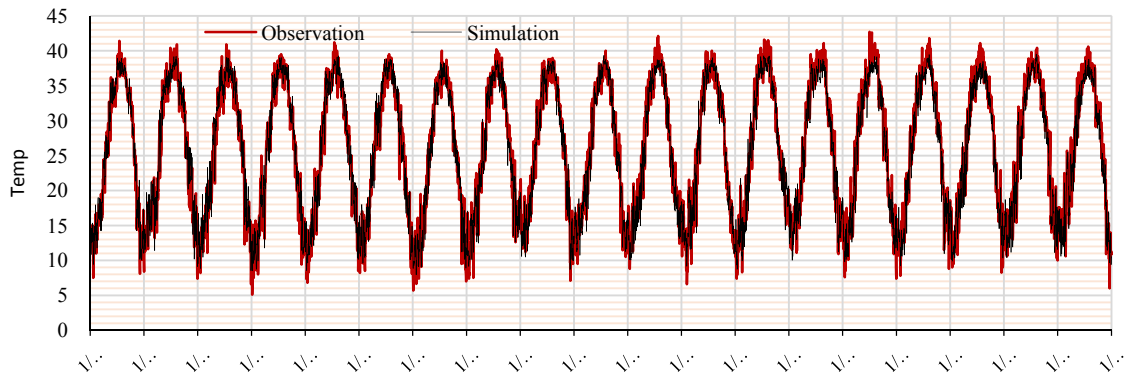
نتایج و بحث

مقایسه مقادیر دمای مشاهداتی و شبیه سازی شده توسط برنامه ریزی ژنتیک در شکل (۳) نشان داده شده است و ضریب همبستگی بین سری داده های مشاهداتی و مدل سازی شده در شکل (۴) ارائه شده است. سری زمانی دماهای مشاهداتی و شبیه سازی شده با استفاده از شبکه عصبی مصنوعی در شکل (۵) نشان داده شده است و ضریب همبستگی آن در شکل (۶) ارائه شده است.

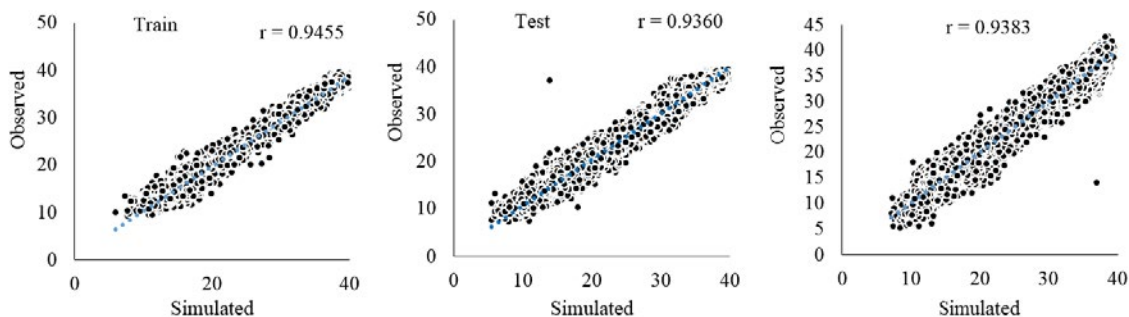
در این بخش به منظور مقایسه عملکرد و دقت مدل های مورد استفاده نتایج حاصل از مدل سازی برای هر روش به صورت نمودار ارائه شده است و سپس ضریب همبستگی آن در مرحله آموزش و تست مورد بررسی قرار گرفته است. در انتها میزان خطای مدل سازی دما برای چهار روش به صورت جدول ارائه شده است.



شکل ۴: ضریب همبستگی بین داده های مشاهداتی و سری شبیه سازی شده به روش برنامه ریزی ژنتیک



شکل ۵: مقایسه مقادیر مشاهداتی و شبیه سازی شده دمای متوسط به روش شبکه عصبی مصنوعی

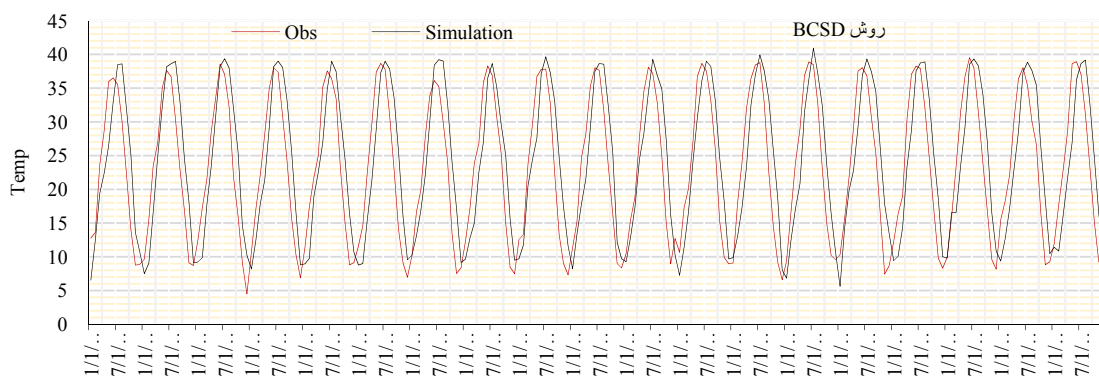


شکل ۶: ضریب همبستگی بین داده های مشاهداتی و سری شبیه سازی شده به روش شبکه عصبی مصنوعی

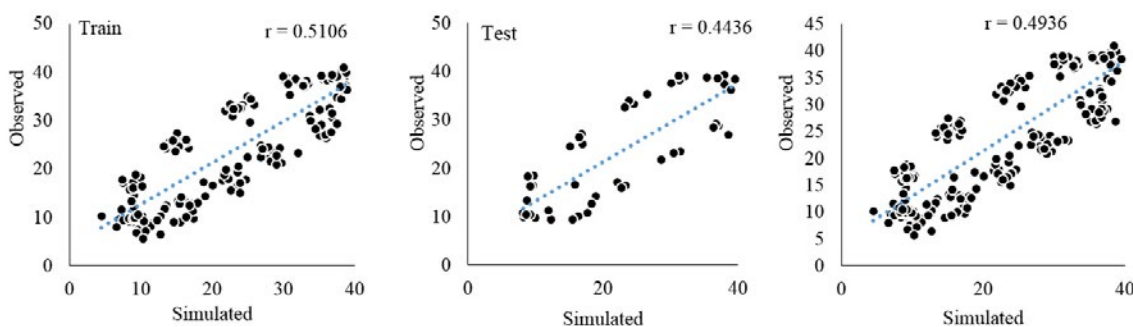
بررسی مقدار خطا

دو معیار ریشه میانگین مربعات خطا (RMSE) که نشان دهنده بزرگی خطا و معیار ناش-ساتکلیف (NSE) نسبت خطای میانگین مربعات به انحراف معیار داده‌های مشاهداتی جهت ارزیابی بهتر مدل‌سازی‌های در جدول (۱) ارائه شده است. مطابق این نتایج، ریشه میانگین

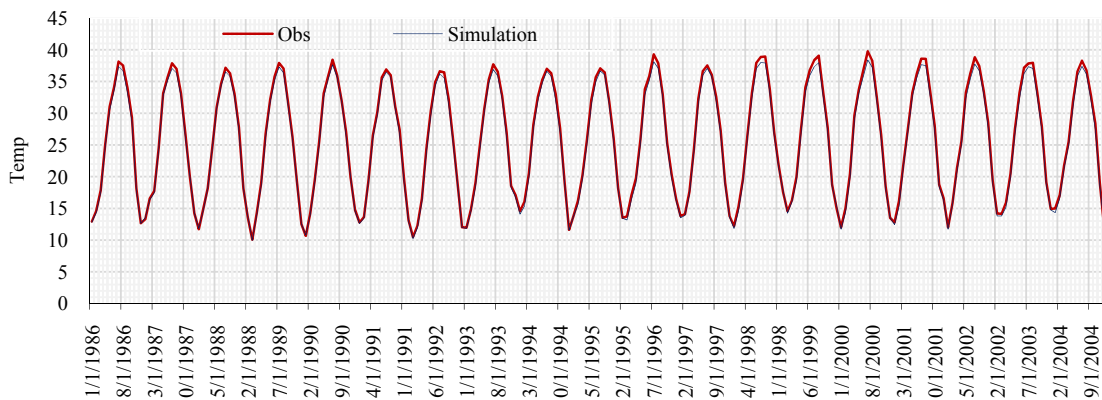
در مقایسه مقادیر مشاهداتی و شبیه‌سازی شده به روش SVM (شکل ۹) مشاهده می‌شود که داده‌های شبیه‌سازی شده نسبت به سایر روش‌ها انطباق بیشتری دارد. به طوری که نمودار شکل (۱۰) نشان دهنده ضریب همبستگی بالا بین مقادیر مشاهداتی و شبیه‌سازی شده و قابلیت نسبی بالای این روش نسبت به سه روش مذکور می‌باشد.



شکل ۷: مقایسه مقادیر مشاهداتی و شبیه‌سازی شده دمای متوسط به روش جداسازی فضایی تصحیح اربیبی



شکل ۸: ضریب همبستگی بین داده‌های مشاهداتی و سری شبیه‌سازی شده به روش جداسازی فضایی تصحیح اربیبی



شکل ۹: مقایسه مقادیر مشاهداتی و شبیه‌سازی شده دمای متوسط به روش ماشین بردار پشتیبان

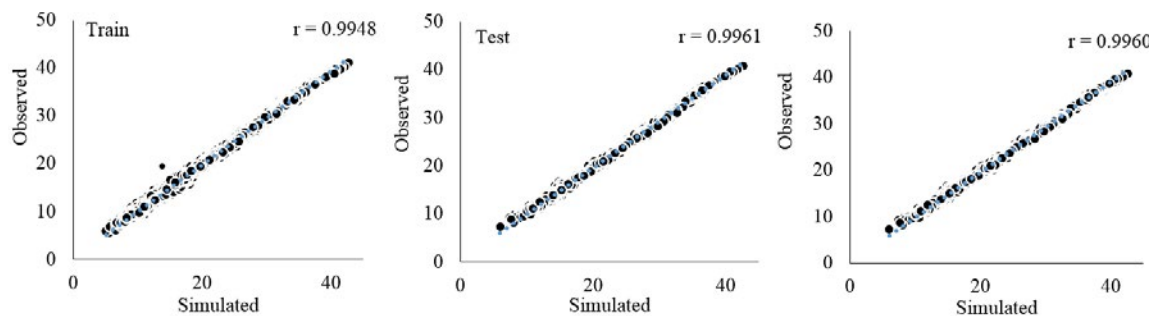
۲ درصد کمتر از مدل SVM است و خطای ناشی از ریشه میانگین مربعات میانگین خطا در مدل ANN ۵۹ درصد بیشتر از مدل SVM است. برتری مدل سازی به روش SVM نسبت به ANN در این پژوهش با نتایج پژوهش (Adamowski 2013) و (Sedighi et al 2016) مطابقت دارد.

نتیجه گیری

مقایسه مدل سازی دمای میانگین روزانه هوا از سری زمانی مشاهداتی در دوره ۲۰۰۴-۱۹۷۱ توسط مدل های برنامه ریزی ژنتیک (GP)، شبکه عصبی مصنوعی (ANN) و جداسازی فضایی تصحیح اربیبی (BCSD) و ماشین بردار پشتیبان (SVM) نشان داد که مدل سازی دما به روش SVM نسبت به سایر روش ها از عملکرد و دقت بالاتری برخوردار است. به طوری که در این روش ضریب همبستگی دماهای مشاهده شده نسبت به مقادیر شبیه سازی شده ۰/۹۹۶۰ برآورد شده است. بزرگی خطا در فرایند شبیه سازی که توسط معیار ریشه میانگین مربعات خطا (RMSE) بررسی شده است در این روش

مربعیات خطا در مدل سازی SVM نسبت به سایر مدل ها کمتر بوده و به ترتیب مدل GP، ANN و BCSD دارای خطای کمتری می باشند. بررسی ریشه میانگین مربعات خطا در مدل سازی BCSD نشان می دهد که مقدار خطا در این روش نسبت به سایر مدل ها به طور معناداری بیشتر می باشد که ناشی از ضعف نسبی این روش در مدل سازی نسبت به سایر روش های مدل سازی می باشد.

مقایسه عملکرد دو مدل GP و ANN با اختلاف ریشه میانگین مربعات خطای ۰/۸ درصد نشان دهنده برتری ضعیف مدل GP نسبت به ANN می باشد. مقایسه عملکرد دو روش GP و SVM نشان می دهد که خطای ناشی از ریشه میانگین مربعات خطا در مدل سازی SVM ۵۸ درصد کمتر از مدل GP بوده است. مقایسه عملکرد این دو مدل که نشان دهنده برتری مدل سازی SVM نسبت به GP می باشد، با نتایج دهقانی و همکاران (۱۳۹۶) و سلگی (۱۳۹۷) مطابقت دارد. مقایسه عملکرد دو مدل ANN و SVM نشان می دهد که در ANN ضریب کارآیی مدل با توجه به معیار ناش-ساتکلایف



شکل ۱۰: ضریب همبستگی بین داده های مشاهداتی و سری شبیه سازی شده به روش ماشین بردار پشتیبان

جدول ۱: ارزیابی خطا (بر حسب درجه سانتی گراد) در مراحل مدل سازی

Evaluation criteria						روش
NSE			RMSE			
در مرحله آموزش	در مرحله آزمایش	کل	مرحله آموزش	مرحله آزمایش	کل	روش
۰/۹۶۸۹	۰/۹۶۹۹	۰/۹۶۹۱	۱/۶۴۰۴	۱/۶۵۵۴	۱/۶۴۴۰	GP
۰/۹۶۸۶	۰/۹۶۸۳	۰/۹۶۸۵	۱/۶۵۱۰	۱/۶۸۱۳	۱/۶۵۷۲	ANN
۰/۶۷۴۷	۰/۵۸۶۳	۰/۶۶۰۶	۶/۰۶۶۰	۶/۶۳۰	۶/۱۷۴۳	BCSD
۰/۹۹۵۲	۰/۹۹۳۰	۰/۹۹۴۷	۰/۶۷۵۶	۰/۶۷۹۵	۰/۶۷۶۹	SVM

۰/۶۷۶۹ درجه سانتی‌گراد می‌باشد که نسبت به سایر روش‌ها این مقدار خطا کمتر می‌باشد. همچنین نسبت خطای میانگین به انحراف معیار مقادیر مشاهداتی که به عنوان ضریب ناش-ساتکلیف (NSE) شناخته می‌شود در این روش برابر ۰/۹۹۴۷ درجه سانتی‌گراد می‌باشد که نزدیک به حالت ایده‌ال (مقدار یک) می‌باشد و نسبت به سایر روش‌ها به حالت ایده‌ال نزدیک‌تر است. همچنین میزان ضریب همبستگی و خطاهای اندازه‌گیری شده برای روش برنامه‌ریزی ژنتیک (GP) و شبکه عصبی مصنوعی (ANN) نشان‌دهنده برتری ضعیف روش برنامه‌ریزی ژنتیک نسبت به شبکه عصبی مصنوعی می‌باشد. انجام شبیه‌سازی‌های مختلف دمای حداقل و حداکثر توسط این دو روش می‌تواند قابلیت هر کدام از این مدل‌ها را بهتر نشان دهد. نتایج حاصل از شبیه‌سازی به روش BCSD با سه روش فوق‌الذکر اختلاف زیادی دارد. به طوری که ضریب همبستگی دماهای شبیه‌سازی شده با سری مشاهداتی ۰/۴۹۳۶ و بزرگی خطای این روش شبیه‌سازی ۶/۱۷۱۴ درجه سانتی‌گراد برآورد شده است که نسبت به سایر روش‌ها مقدار بیشتری می‌باشد. مقدار ضریب ناش-ساتکلیف این روش به ۰/۶۶۰۶ درجه سانتی‌گراد محاسبه شده است که نسبت به سایر روش‌های مدل‌سازی با حالت ایده‌ال اختلاف بیشتری دارد. بنابراین انجام شبیه‌سازی دما و تغییرات دمای هوا در دوره‌های آینده توسط SVM می‌تواند مقادیر را با خطای کمتر و دقت بیشتر پیش‌بینی کرده و از این طریق برنامه‌ریزی‌های مربوطه و به کارگیری تمهیدات لازم جهت مدیریت این تغییرات را با ریسک کمتر مواجه سازد.

منابع

- اسکندری، ع.، نوری، ر.، معراجی، ع. کیا قادری، ا. ۱۳۹۱. توسعه مدلی مناسب بر مبنای شبکه عصبی مصنوعی و ماشین بردار پشتیبان برای پیش‌بینی به هنگام اکسیژن‌خواهی بیوشیمیایی ۵ روزه، مجله محیط‌شناسی، ۳۸(۶۱): ۷۱-۸۲.
- دهقانی، ر.، یونسی، ح.، ترابی پوده، ح. ۱۳۹۶. مقایسه عملکرد مدل‌های ماشین بردار پشتیبان، برنامه‌ریزی بیان ژن و سلگی، ا.، زراعی، ح.، شهنی‌دارابی، م.، علی‌دادی، س. ۱۳۹۷. پیش‌بینی بارش ماهانه با استفاده از مدل‌های برنامه‌ریزی بیان ژن و ماشین بردار پشتیبان، نشریه تحقیقات کاربردی علوم جغرافیایی، ۱۸(۵۰): ۹۱-۱۰۳.
- شاهی‌نژاد، ب.، دهقانی، ر. ۱۳۹۷. مقایسه مدل‌های شبکه عصبی موجک، ماشین بردار پشتیبان و برنامه‌ریزی بیان ژن در تخمین میزان اکسیژن محلول در رودخانه‌ها. نشریه تحقیقات منابع آب ایران، ۱۴(۳): ۲۶۵-۲۷۷.
- علی‌پور، ع.، موسوی، ح.، خلیلیان، صادق. ۱۳۹۳. ارزیابی هزینه انتشار گاز گلخانه‌ای کربن دی‌اکسید حاصل از توسعه بخش کشاورزی ایران. مجله اقتصاد کشاورزی، ۸(۱): ۸۱-۶۳.
- کدخدا حسینی، م.، میرعباسی، ر.، نودری، ح.، رستمی، ع. ۱۳۹۷. ارزیابی عملکرد رهیافت‌های برنامه‌ریزی ژنتیک و ماشین بردار پشتیبان در بازسازی داده‌های گم‌شده بارش. نشریه هواشناسی کشاورزی، ۶(۱): ۴۹-۴۱.
- منهاج، م.ب. ۱۳۹۳. هوش محاسباتی، جلد اول: مبانی شبکه‌های عصبی. انتشارات دانشگاه صنعتی امیرکبیر، ۷۱۶ ص.
- نادری، س.، گودرزی، م.، قدمی‌دهنو، م. ۱۳۹۶. تاثیر تغییرات اقلیمی بر پارامترهای آب و هوایی حوضه سیمار. مجله مهندسی و مدیریت آبخیز، ۱۱(۳۹): ۶۹-۷۶.
- Adamowski J., and Prasher S.O. 2012. Comparison of machine learning methods for runoff forecasting in mountainous watersheds with limited data. *Journal of Water and Land Development*, 17(7-8): 89-97.
- Ahmed, K.F., Wang, G., Silander, J., Wilson, A.M., Allen, J.M., Horton, R., and Anyah, R. 2013. Statistical downscaling and bias correction of climate model outputs for climate change impact assessment in the U.S. northeast. *J. Global Planet Change*, 100: 320-332.
- Anagu, I., Joachim, I., Jens, U., and Thilo, S. 2009. Estimation of heavy metal sorption in German soils using artificial neural networks. *Geoderma* 152: 104-112.
- Basak, D., Pal, S., Patranabis, D.CH. 2007. Support Vector Regression. *Neural Information Processing – Letters and Reviews*. 11(10):203-224.

- Water resources management, 26(2): 457–474.
- Koza J.R. 1992. Genetic Programming: on the programming of computers by means of natural selection. Cambridge, MA: MIT Press.
- Nash, J.E., and Sutcliffe, J.V. 1970. River flow forecasting through conceptual models part I-A discussion of principles. *Journal of Hydrology*, 10. 3: 282–290.
- Salinger, M. J., Stiger, C. J. Dasc, H. P. 2000. Agro meteorological adaption strategies to increasing climate variability and climate change. *Journal of Agricultural and Forest Meteorology*, 103:167-184.
- Sedighi, F., Vafakhah, M., and Javadi, M. R. 2016. Rainfall–Runoff modeling using support vector machine in snow-affected watershed. *Arabian Journal for Science and Engineering*. 41: 10. 4065-4076.
- Tezel, G. and Buyukyildiz, M. (2015). Monthly evaporation forecasting using artificial neural networks and support vector machines, *Theoretical and Applied Climatology*.
- Thiery, F., Grieu, S., Traoré, A. 2008. Integration of neural networks in a geographical information system for the monitoring of a catchment area. *Mathematics and Computers in Simulation* 76 : 388–397.
- Vanuytrecht, E., Raes, D., Willems, P., Geerts, S. 2012. Quantifying field-scale of elevated carbon dioxide concentration on crops. *Climate Research*, 54: 35-47.
- Vapnik, V. 1984. Estimation of dependences based on empirical data. Springer-Verlag, 400 p.
- Vapnik, V., Chervonenkis, A. 1974. Theory of pattern recognition. Nauka, Moscow. 353 p.
- Borelli, A., De Falco, I., Della, CA., Nicodemi, M., Trautteur, G. 2006. Performance of genetic programming to extract the trend in noisy data series. *scientific journals of physics*, 370: 104-108.
- Chen, L. 2003. study of applying genetic programming to reservoir trophic state evaluation using remote sensor data, *International Journal of Remote Sensing*, 24(11): 2265-2275.
- Feng, H., and Liu, Y. 2015. Combined effects of precipitation and air temperature on soil moisture in different land covers in a humid basin, *Journal of Hydrology*, 53:1, 1129–1140.
- Ferreira, C. 2001. Gene Expression Programming: A New Adaptive Algorithm for Solving Problems. *Complex Systems*, forthcoming.
- Ghorbani, M.A., Ahmadzadeh, H., Isazadeh, M. and Terzi, O. 2016. A comparative study of artificial neural network (MLP, RBF) and support vector machine models for river flow prediction. *Environmental Earth Sciences* 1-14 :(6)75 .
- Hyndman, R. J., Koehler, A. B. 2006. Another look at measures of forecast accuracy. *International Journal of Forecasting*. 22(4): 679–688.
- IPCC, 2001. Climate change: The scientific Basis, Contribution of working Group to the Third Assessment Report of the intergovernmental Panell on climate change. Cambridge Univ. Press, New York, NY, USA, 881 P.
- Kisi, O., Moghaddam Nia, A., Ghafari Gosheh, M., Jamalizadeh Tajabadi, M.R. and Ahmadi, A. 2012. Intermittent Streamflow Forecasting by Using Several Data Driven Techniques.

Performance comparison of GP, ANN, BCSD and SVM models for temperature simulation

S.M. Tabatabaei ^{1*}, M. Nazeri Tahroudi ¹, M. Dastourani ²

¹Faculty of Agriculture, University of Birjand, Birjand, Iran ² Assistant Professor, Department of Water Science and Engineering, Faculty of Agriculture, University of Birjand, Birjand, Iran

*Corresponding Author Email: tabatabaei1984@yahoo.com

Received: 9 March 2018, accepted: 11 June 2018

ABSTRACT

In this study, GP, ANN, BCSD and SVM were used to evaluate the performance of simulation models. The modeling based on the large scale data of atmospheric general circulation and the mean daily temperature of the Ahwaz synoptic station was conducted during the period of 1971-2004, and the evaluation of each model was based on the correlation coefficient and modeling error between simulated and observational data. The results of modeling indicated that the correlation coefficient between observation and simulation data in the SVM model is more than in other models, and its value is 0.9960. The correlation coefficient for the GP, ANN and BCSD models is equal to 0.9393, 0.9384, and 0.4936 respectively. Moreover, the results of the evaluation of the simulation error were calculated using the RMSE and NSE criteria for SVM to be 0.677 and 0.955 degrees Celsius respectively. Similarly, these values are 1.644 and 0.969 for GP, 1.657 and 0.968 for ANN, and 6.174 and 0.661 for BCSD. Therefore, SVM has better performance in modeling the mean daily temperature than other methods, and modeling the mean daily temperature by the BCSD method is less accurate than other methods. The GP method has a weak advantage over ANN, and it is recommended to use the minimum and maximum temperatures for the more precise performance of these two models.

Keywords: Genetic programming, Climate Change, Modeling, Support Vector Machines

HOW TO CITE THIS ARTICLE

Tabatabaei, S.M.; Nazeri-tahroudi, M.; Dastourani, M., (2018). Performance comparison of GP, ANN, BCSD and SVM models for temperature simulation . *J. Meteorol. Atmos. Sci.*, 1(1): 53-64

COPYRIGHTS

Copyright for this article is retained by the author(s), with publication rights granted to the JMAS Journal. This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>).

