Archive of SID

فصلنامه علمی-پژوهشی تحقیقات مرتع و بیابان ایران
جلد ۲۲، شماره ۲، صفحه ۲۵۰-۲۴۰ (۱۳۹۴)

تعیین مناسب‌ترین روش ترکیب ورودی شبکه عصبی مصنوعی به‌منظور تعیین عوامل باد بر پیش‌بینی پدیده طوفان گرد و غبار (مطالعه موردی: استان یزد)

محسن یوسفی[۱*] و لیلا کاشی زنوزی[۲]

۱*- نویسنده مسئول، دانشجوی کارشناسی ارشد آبخیزداری، دانشکده منابع طبیعی و کویرشناسی، دانشگاه یزد، ایران،
پست الکترونیک: mohsenyosefi67@gmail.com
۲- کارشناس‌ارشد پژوهشی، بخش تحقیقات بیابان، مؤسسه تحقیقات جنگلها و مراتع کشور، سازمان تحقیقات، آموزش و ترویج کشاورزی، تهران، ایران

تاریخ دریافت: ۹۱/۱۲/۱ تاریخ پذیرش: ۹۲/۵/۴

چکیده

هدف از این مطالعه تعیین برخی عوامل تأثیر گذار بر پدیده طوفان گردوغبار با استفاده از روش‌های مختلف است. به‌منظور تعیین مناسب‌ترین ترکیب ورودی، از روش‌های کاهش متغیر از قبیل تحلیل عاملی (حداکثر احتمال، تجزیه مؤلفه‌های اصلی)، آزمون گاما و رگرسیون چند متغیره استفاده شد. هر کدام از روش‌های مذکور ترکیب متفاوتی را ارائه نمودند که هر کدام از این ترکیب‌ها در مدل شبکه عصبی پیشخور پس انتشار با توابع آموزشی لورنبرگ مارکوآت استفاده شد که رگرسیون گام به گام با R^2=۰/۸۷ و RMSE=۰/۰۴ مناسب‌ترین ترکیب را برای مدل شبکه عصبی معرفی نمود همچنین داده‌ها را به‌صورت ماهانه و فصلی با استفاده از مناسب‌ترین ورودی به شبکه اعمال شد و شبیه‌سازی پدیده طوفان گردوغبار در فصل‌های تابستان و بهار و در ماه‌های اردیبهشت، فروردین، خرداد، تیر، شهریور و مرداد با شاخص‌های آماری ضریب همبستگی بالاتر و میانگین مربعات خطای پایین‌تر بدلیل پراکنش مناسب داده‌های طوفان گرد و غبار انجام شد. نتایج تحقیق نشان داد که بر اساس روش‌های به‌کار رفته، بیشترین تأثیر بر پدیده طوفان گرد و غبار را در استان یزد، عوامل سرعت باد غالب، دید افقی، تداوم و میانگین سرعت باد به عهده دارند.

واژه‌های کلیدی: شبکه عصبی مصنوعی، طوفان گرد و غبار، رگرسیون گام به گام، آزمون گاما، تجزیه مؤلفه‌های اصلی، تجزیه عاملی.

مقدمه

استان یزد ۱۹ درصد بیابان‌های کشور را در خود جای داده و دومین استان بیابانی کشور محسوب می‌شود. طوفان‌های گرد و غبار حوادث طبیعی هستند که در مناطق خشک، نیمه‌خشک و بیابانی دنیا به فراوانی رخ می‌دهند (Song، ۲۰۰۴). طوفان پدیده‌ای نامنظم از نظر زمانی و مکانی بشمار می‌آید که بیشتر در اواخر زمستان و اوایل بهار اتفاق می‌افتد و همواره با خسارت‌های مالی، اقتصادی و گاهی جانی توأم است. وجود ماسه‌های روان که یکی از آشکارترین آثار فرسایش بادی است، در این استان فراوان دیده می‌شود. ماسه‌های روان می‌توانند تا شعاع نسبتاً وسیعی از اطراف خود را تحت تأثیر قرار دهند و شهرهای این استان را ساعت‌ها در تاریکی فرو برند و سبب راه‌بندان شوند. کانون‌های بحرانی حساس به فرسایش بادی منشأ وقوع طوفان‌های شن در استان یزد محسوب می‌شوند. ۶۵۹ هزار هکتار از عرصه‌های بیابانی در ۱۹ نقطه از استان یزد به‌عنوان کانون‌های بحرانی حساس به فرسایش بادی مورد شناسایی قرار گرفته که با هر بار وقوع طوفان مقادیر زیادی ذرات خاک از این کانون‌ها بلند شده و در آسمان شهرها و نقاط سکونت‌گاهی استان دیده می‌شود. به دلیل موقعیت

www.SID.ir

Archive of SID

جغرافیایی و شرایط اقلیمی، این منطقه همواره در معرض بادهای شدید و طوفان‌های گرد و غبار قرار دارد. از این جهت شناخت رژیم بادهای شدید و طوفان‌های منطقه یزد به‌منظور کاهش آثار مخرب این پدیده به‌ویژه در امر تثبیت ماسه‌های روان ضروریست. به‌طوری‌که وزش بادهای غالب از سمت شمال‌غرب و غرب است. بیش از ۵۰ درصد از بادهای شدید و طوفان‌های منطقه در ماه‌های اردیبهشت و فروردین رخ می‌دهند. علت اصلی این پدیده وزش بادهای شدیدی است که بدنبال تغییرات سریع فشار و دمای هو ا در این موقع از سال رخ می‌دهد.

پیش‌بینی موفق این رخدادها می‌تواند کمک زیادی به بهبود زندگی در این مناطق باشد. با هشدارهای لازم قبل از وقوع طوفان گرد و غبار می‌توان شهروندان را از خطرات بهداشتی وارده از طرف این طوفان‌ها از قبیل تنگی نفس آگاه کرد. همچنین در بخش ترابری هوایی و زمینی کاهش دید ناشی از وقوع این طوفان‌ها خسارت‌آفرین است که مقابله با این خسارت‌ها نیازمند پیش‌بینی این رخدادها می‌باشد.

در سطح جهان بررسی‌های زیادی در مورد روز طوفانی و طوفان‌های گرد و غبار انجام شده که در این مقاله به برخی از آنها اشاره می‌شود.

Chen و همکاران (۱۹۹۶) در این مورد مطالعاتی را در شمال چین انجام داده و بیان کردند که با افزایش رطوبت خاک میزان طوفان‌های شدید و فرسایش بادی کاهش پیدا می‌کند. Feng وEndo Chang (۲۰۰۲) در زمینه تأثیر طوفان‌های خاک را بر روی سلامتی و تنفس بررسی کردند و تغییرات تمرکز غبار در طول طوفان‌های خاک در تایوان را مورد مطالعه قرار داده و نشان دادند رابطه معنی‌داری میان آنها وجود دارد. Orlovsky و همکاران (۲۰۰۵) طوفان‌های خاک را در ترکمنستان مورد مطالعه قرار دادند و توزیع فضایی، فراوانی و تغییرات فصلی این طوفان‌ها را بررسی کرده و بیان کردند که بیشترین تعداد روز طوفانی در فصل بهار و در منطقه بیابانی قره‌قوم وجود دارد. Jamalizadeh و همکاران (۲۰۰۸) در تحقیقی تحت عنوان پیش‌بینی وقوع طوفان گرد و غبار با استفاده از روش شبکه عصبی مصنوعی، به این نتایج دست یافتند که این روش در پیش‌بینی کوتاه‌مدت وقوع طوفان‌ها موفقیت بیشتری را نشان می‌دهد (d=۰/۹۶)، اگرچه با بیشتر شدن زمان پیش‌بینی، از دقت نتایج کاسته می‌شود (d=۰/۹۵)؛ در حالی‌که در پیش‌بینی میزان دید موفقیت کمتری بدست آمد (d=۰/۸۵). بنابراین بنظر می‌رسد با شناخت بهتر فرایند این طوفان‌ها بتوان پیش‌بینی‌های دقیق‌تری را با استفاده از این شبکه‌ها انجام داد. Huang و همکاران (۲۰۰۶) از روش شبکه عصبی مصنوعی برای پیش‌بینی طوفان گرد و خاک در شمال‌غرب چین استفاده کردند و به این نتایج رسیدند که به‌طور متوسط ۷۱/۶ و ۶۸/۲ درصد طوفان‌ها را پیش‌بینی نمایند. طبق یافته‌هایRashki و همکاران (۲۰۱۲) در پژوهشی تحت عنوان طوفان‌های گرد و خاک و بارگذاری افقی گرد و غبار در منطقه سیستان، بر اساس آنالیز AQI ۶۱٪ روزها دارای هوای آرام و ۳۰/۱٪ دارای خطر بودند.

طوفان گرد و غبار در منابع مختلف تعاریف متعدد و تا حدودی نامشخص دارد. اما مهمترین متغیر در تشخیص وقوع طوفان گرد و غبار، وجود ذرات گرد و غبار در هواست. در هنگام ثبت داده‌ها در ایستگاه‌های هواشناسی سینوپتیک، هوایی که غبار داشته یا طوفان شن یا گرد و غبار رخ داده باشد به چند دسته تقسیم می‌شود که مبنای این تقسیم‌بندی نیز بیشتر مقدار دید افقی است (Westwell, 1999). در این تحقیق به پیش‌بینی مواردی از این طوفان‌ها پرداخته‌ایم که دید افقی در آنها به کمتر از ۲۰۰۰ متر می‌رسد.

هدف از این مطالعه بدست آوردن یک روش مناسب برای تعیین مناسب‌ترین ترکیب ورودی برای مدل شبکه عصبی مصنوعی و تعیین عوامل تأثیرگذار بر پدیده طوفان گرد و غبار است و مهمترین فصل و ماه در طوفان گرد و غبار انتخاب شد.

مواد و روش‌ها

موقعیت منطقه مورد مطالعه

استان یزد با مساحت حدود ۱۳۱۵۷۵ کیلومترمربع در قسمت مرکزی فلات ایران قرار دارد. این استان در $29^{\circ}35'$

www.SID.ir

Archive of SID

تا ′۰۷ °۳۵ عرض شمالی و ′۵۰ °۵۲ تا ′۱۶ °۵۸ طول شرقی واقع شده است و دربرگیرنده نامناسب‌ترین عوامل طبیعی غالب بر فلات مرکزی ایران است. آب و هوای استان یزد بعلت قرار داشتن بر روی کمربند خشک جهانی دارای زمستان‌های سرد و نسبتاً مرطوب و تابستان‌های گرم و طولانی و خشک است. شدت وزش باد در استان بعلت لخت بودن دشت‌ها و کوهستان‌ها زیاد می‌باشد. جهت وزش باد در شش ماه اول سال شمال‌غربی و در چهار ماه آبان تا بهمن جنوب‌شرقی می‌باشد. وزش باد غالب در دو ماه اسفند و مهر از غرب می‌باشد. سرعت باد به‌صورت طوفان‌های سهمگین شنی می‌تواند تا ۹۰ کیلومتر در ساعت برسد و حتی این سرعت در یزد تا ۱۲۰ کیلومتر نیز ثبت شده است. بر اساس آمار موجود در ایستگاه سینوپتیک یزد، میانگین سرعت باد در این ایستگاه ۲/۵۵ متر بر ثانیه است و میانگین سرعت باد غالب ۲۱ متر بر ثانیه بدست آمده است.

داده‌های مورد استفاده

در این مطالعه، داده‌های سینوپتیک ایستگاه یزد در بازه زمانی ۱۹۵۳– ۲۰۰۵ به‌صورت ماهانه مورد استفاده قرار گرفت. داده‌ها پس از اخذ از سازمان هواشناسی کشور، مورد پردازش قرار گرفته و داده‌های ناقص بازسازی شدند. داده‌های ورودی عبارتنداز: طوفان تندر، بزرگی باد، پیوستگی باد، دید افقی، سریع‌ترین سرعت باد، میانگین سرعت باد، سرعت باد غالب و طوفان گرد وخاک به‌عنوان خروجی مدل معرفی شد. البته لازم به ذکر است که می‌توان از متغیرهای بیشتری به‌عنوان ورودی استفاده کرد ولی بدلیل نامناسب بودن برای مدل شبکه عصبی مصنوعی (جواب نامناسب مدل) مورد استفاده قرار گرفته نشد. البته داده‌های پاییز و زمستان بدلیل اینکه طوفان گرد و غبار در ماه‌های مذکور کمتر اتفاق افتاده و بیشتر داده‌ها صفر می‌باشد در جواب مدل تأثیر منفی می‌گذارند. بنابراین به‌منظور جلوگیری از این مشکل داده‌ها در فصول مذکور حذف شدند.

در این مطالعه از مدل شبکه عصبی پیشخور پس انتشار خطا با الگوریتم آموزشی لورنبرگ مارکوآت (Levenberg-Marquardt:LM) با تابع انتقال سیگموئید (این تابع انتقال در شبکه‌های پس انتشار) استفاده گردید. این تابع انتقال مقادیر ورودی را در محدوده ∞− تا ∞+ دریافت کرده و خروجی بین یک و صفر را تولید می‌نماید. این تابع متداول‌ترین تابع انتقال بوده و شکلی مانند S دارد. این تابع با آنکه غیرخطی می‌باشد اما رفتار آن مورد توجه و علاقه می‌باشد.

مدل شبکه عصبی مصنوعی

شبکه‌های عصبی مصنوعی الگویی برای پردازش اطلاعات هستند که با تقلید از شبکه عصبی بیولوژیکی مثل مغز انسان ساخته شده‌اند. عنصر کلیدی این الگو ساختار جدید سیستم پردازش اطلاعات آن است که از تعداد زیادی عنصر نرون با ارتباطات قوی داخلی که هماهنگ با هم برای حل مسائل مخصوص کار می‌کنند، تشکیل شده است (Bauer *et al.*,2007).

روش شبکه عصبی مصنوعی از توزیع آماری داده‌ها مستقل است و به متغیرهای آماری مخصوصی نیاز ندارد، علاوه بر آن روش شبکه عصبی مصنوعی برای آنالیز صحت داده‌ها به بررسی‌های کمتری در قیاس با روش‌های آماری نیاز دارد (Caniani *et al.*, 2008 و Lee *et al.*, 2006). شبکه‌های عصبی مصنوعی ازجمله مدل‌های شبیه‌سازی است که می‌تواند با دقتی در خور توجه واقعیت‌های موجود را به تصویر بکشد و راهکار مناسبی برای برآورد و مدل‌سازی باشد (Hsu *et al.*,1995).

الگوریتم لونبرگ مارکوآرت

این روش سعی در کاهش محاسبات با استفاده از عدم محاسبه ماتریس هیسین دارد. بسیار سریع‌تر از بقیه الگوریتم‌ها عمل می‌کند. اشکال عمده روش لونبرگ مارکوآرت نیاز آن به نگهداری ماتریس‌های حجیم در حافظه است و این مسئله نیاز به فضای زیادی دارد (کیا، ۱۳۸۹).

سنجش میزان کارآیی مدل

از تجزیه و تحلیل خطاهای باقیمانده و تفاوت‌های بین مقادیر اندازه‌گیری شده و برآورد شده می‌توان در ارزیابی

www.SID.ir

Archive of SID

کارآیی مدل بهره برد. به‌طوری‌که شمار زیادی از ملاک‌ها برای ارزیابی کارآیی مدل وجود دارند (خداوردی‌لو و همکاران، ۱۳۸۳). در پایان، نتایج حاصل از مدل‌ها (خروجی‌ها) و رخدادهای واقعی طوفان با همدیگر مقایسه شده و از شاخص معیارهای همبستگی (R^2)، و مجذور میانگین مربعات خطا (RMSE) برای نشان‌دادن عملکرد مدل‌ها استفاده شده است.

رابطه (۱)
$$R^2 = \frac{\sum_{k-1}^{k} X_k Y_k}{\sqrt{\sum_{k-1}^{k} X_k^2 \sum Y_k^2}}$$

رابطه (۲)
$$RMSE = \sqrt{\frac{\sum_{k-1}^{k} (X_k - Y_k)^2}{K}}$$

به‌طوری‌که در این روابط X_k مقدار مشاهداتی، Y_k مقدار برآوردشده، o و p به‌ترتیب مقادیر مشاهده شده و شبیه‌سازی شده، σ و p به‌ترتیب میانگین مقادیر مشاهده شده و شبیه‌سازی، n و K تعداد نمونه‌ها می‌باشد.

اگر فرض شود a عامل ورودی بر وقوع پدیده‌ای مؤثر باشد، باید تعداد 2^a-1 ترکیب معنی‌دار از عوامل ورودی بوجود بیاید که برای مدل‌سازی این پدیده با استفاده از ANNs بررسی تک تک ترکیبات ایجاد شده برای یافتن بهترین ترکیب کاری بسیار وقت‌گیر و خسته‌کننده است. بنابراین هنگامی که عوامل مؤثر بر پدیده‌ای به‌طور قابل ملاحظه‌ای زیاد باشد با استفاده از روش‌های مختلف می‌توان ترتیب میزان اهمیت عوامل ورودی و بهترین ترکیب از میان تمام ترکیب‌های ممکن را بدست آورد.

رگرسیون چندمتغیره

وقتی که تعداد عوامل ورودی افزایش می‌یابد ارزیابی کلیه روش‌های رگرسیون چندمتغیره به حجم محاسبات زیادی نیاز دارد. بنابراین روش‌های مختلفی ایجاد شده است که فقط تعداد کمی از مدل‌های رگرسیون چندمتغیره دارای زیر مجموعه‌ای از متغیرها را در یک زمان بررسی می‌کنند. روش‌های رگرسیون چندمتغیره را می‌توان در سه دسته کلی گزینش پیش‌رونده (FS)، حذف پس‌رونده و رگرسیون گام به گام که ترکیبی از روش‌های پیشین است، تقسیم نمود. در روش رگرسیون پیشرونده که براساس مدل رگرسیون خطی بنا نهاده شده است، برای قضاوت اینکه در هر مرحله آیا یک متغیر مستقل جدید باید به مدل افزوده شود یا خیر، یک برابر α انتخاب می‌کنند، که در این مطالعه سطح معنی‌داری α سطح ۰/۰۵ انتخاب شده است. به بیان دیگر مدل رگرسیون ساده را برای هریک از عوامل مستقل، برازش داده و برای هر مدل رگرسیون ساده آماری مقدار آماره F را محاسبه می‌کنند. آن متغیر مستقلی که دارای مقدار F بزرگتری است، برای افزودن به مدل انتخاب می‌شود. اگر F مربوط به این متغیر مستقل انتخاب شده از $F\alpha$ (1,n-2) بزرگتر باشد، آنگاه متغیر مستقل مربوطه به مدل افزوده می‌شود، در غیر این صورت افزودن این متغیر مستقل به مدل چندان مفید نخواهد بود. سرانجام مجموعه‌ای که با افزودن عامل مستقل دیگر در سطح α افزایش معنی‌داری در آماره F ایجاد نکند، به‌عنوان بهترین ترکیب ورودی برای مدل‌سازی انتخاب می‌گردد.

آزمون گاما

آزمون گاما یک ابزار مدل‌سازی غیرخطی است که به کمک آن می‌توان ترکیب مناسب از عوامل ورودی برای مدل‌سازی داده‌های خروجی و ایجاد یک مدل هموار را بررسی نمود. با فرض آن که مجموعه‌ای از داده‌های ورودی x_i و خروجی y_i مشاهده شده از پدیده‌ای را به‌صورت رابطه ۳ در اختیار داشته باشیم.

www.SID.ir

Archive of SID

رابطه (۳)
$$\{(x_i, y_i), 1 \le i \le M\}$$

در صورتی‌که با توجه به مجموعه ورودی xi، خروجی y_i بدست آید و M نشان‌دهنده تعداد سری مجموعه مشاهده شده از پدیده مورد بررسی باشد، آنگاه آزمون گاما با فرض عدم قطعیت و پذیرش مقدار خطا در بدست آوردن خروجی از روی ورودی بدلیل پیچیدگی و غیرخطی بودن پدیده‌های مدل‌سازی این خطا را به‌صورت رابطه ۴ بین مجموعه داده‌های ورودی و خروجی با تابع f نشان می‌دهد.

رابطه (۴)
$$y = f(x_1, x_2, \ldots, x_m) + r$$

که در آن f نمایانگر تابع همواری است که برای مدل‌سازی داده‌ها استفاده می‌شود و r نشان‌دهنده متغیر تصادفی است که برای نمایش خطا به‌کار می‌رود. بر اساس N[i,k] می‌باشد که در آن p بیانگر نزدیک‌ترین همسایگی است. مقادیر این آزمون برای مجموعه داده‌های ورودی و خروجی بر اساس روابط ۵ و ۶ بدست می‌آید که {...} فاصله اقلیدسی است.

رابطه (۵)
$$\delta_M(k) = \frac{1}{M}\sum_{i=1}^{M}\left|x_{N(i,k)} - x_i\right|^2 \qquad 1 \le k \le p$$

رابطه (۶)
$$\gamma_M(k) = \frac{1}{2M}\sum_{i=1}^{M}\left|y_{N(i,k)} - y_i\right|^2 \qquad 1 \le k \le p$$

با ایجاد رابطه رگرسیون خطی بین P مجموعه (γM (k (δ M (k))، مقدار آماره گاما برابر عرض از مبدأ خط رگرسیون ایجاد شده است که معادله این خط در رابطه ۷ آورده شده است.

رابطه (۷)
$$\gamma = A\delta + \Gamma$$

یکی دیگر از معیارهای مهم که با استفاده از این آزمون می‌توان بدست آورد معیار بدون بعد V_{ratio} است که دارای مقادیری بین بازه صفر و یک است و هرچه این مقدار به صفر نزدیکتر باشد نمایانگر دقت بالای مدل برای یافتن خروجی‌های مطلوب از ورودی‌هاست. در واقع اگر مقدار V_{ratio} از عدد یک کم شود مقدار ضریب تبیین را نشان می‌دهد.

تحلیل عاملی (FA)

تجزیه و تحلیل عاملی یکی از روش‌های چندمتغیره آماری است که با آن می‌توان تعداد زیادی از متغیرها را به چند عامل کاهش داد و به این ترتیب خلاصه‌ای از مهمترین داده‌های اصلی را مشخص نمود. در حقیقت، هدف تشخیص

www.SID.ir

Archive of SID

این عامل‌های مشاهده‌ناپذیر بر پایه مجموعه‌ای از متغیرهای قابل مشاهده است. عامل، متغیر جدیدی است که از طریق ترکیب خطی نمره‌های اصلی متغیرهای مشاهده شده بر پایه فرمول زیر برآورد می‌شود:

$$Fj=\sum WjiXi=Wj1X1+Wj2X2+\ldots+WjpXp \qquad \text{رابطه (۸)}$$

که در آن Wها بیانگر ضرایب نمره عاملی و P معرف تعداد متغیرهاست.

امید این است که با تعداد کمی از این عامل‌ها (یعنی ترکیب‌های خطی نمره‌های اصلی متغیرهای مشاهده شده)، بتوان تقریباً همه اطلاعاتی را که توسط مجموعه بزرگتری از متغیرها بدست می‌آید، دربرگرفته و در نتیجه توصیف ویژگی‌های هر مشاهده را ساده کرد.

تجزیه مؤلفه‌های اصلی (PCA: principal analysis components)

در تجزیه مؤلفه‌های اصلی کلیه متغیرها به‌طور مساوی در نظر گرفته می‌شوند. در این روش بر خلاف روش رگرسیون چندگانه متغیرها به دو گروه متغیرهای مستقل و وابسته تقسیم نمی‌شوند، بلکه هر مؤلفه اصلی مقداری از واریانس کل را تبیین می‌کند. بنابراین اولین مؤلفه اصلی حاوی بیشترین اطلاعات و دارای بالاترین واریانس است و آخرین مؤلفه دارای کمترین مقدار واریانس می‌باشد. در تعیین مهمترین مؤلفه‌های اصلی، تعیین مقدار ویژه و بردارهای ویژه حائز اهمیت هستند، از این‌رو بشرح مختصری در مورد آنها می‌پردازیم:

مقادیر ویژه و بردارهای ویژه: مقدار ویژه به زبان ساده عبارت است از تعیین واریانس و انحراف معیار در ماتریس. از آنجا که اولین مرحله در تجزیه مؤلفه‌های اصلی ایجاد ماتریس همبستگی است، از این‌رو گام بعدی پس از تشکیل ماتریس اولیه، تعیین مقادیر ویژه و بردارهای ویژه می‌باشد. محاسبات این بردارها از طریق محاسبات تکراری انجام می‌شود.

نتایج

نتایج آزمون رگرسیون پیش‌رونده

بر اساس نتایج این آزمون در مرحله نخست سرعت باد غالب مهمترین عاملی است که برای ورود به مدل انتخاب می‌شود. زیرا مطابق با نتایج جدول ۱ که در آن مقادیر ضریب همبستگی ساده (دو به دو) عوامل هواشناسی آمده است، عامل سرعت باد غالب دارای بیشترین مقدار ضریب همبستگی با طوفان گرد و خاک (روز طوفانی) به مقدار ۰/۸۶ می‌باشد. بنابراین وقتی به مدل اضافه می‌شود مقدار ضریب تبیین را به ۷۳/۴۷ می‌رساند. در گام بعدی عامل دید افقی با بیشترین مقدارF نسبت به سایر عوامل مستقل باقی مانده، به مدل اضافه می‌شود که باعث افزایش این ضریب به عدد ۸۰/۳۱ می‌شود. سپس از میان مجموعه عوامل باقیمانده عاملی که می‌تواند با بیشترین مقدارF و معنی‌داری در سطح α برابر ۰/۰۵ به مدل اضافه شود، پیوستگی باد است که ضریب تبیین مدل را به ۸۵/۲۷ ارتقاء می‌دهد و به همین ترتیب عواملی که در گام‌های بعدی می‌توانند در سطح α باعث ایجاد معنی‌داری در مقدارF شوند، عامل میانگین سرعت می‌باشد و در نهایت اضافه نمودن عوامل طوفان تندر، بزرگی باد و سریع‌ترین سرعت باد نمی‌تواند در سطح ۰/۰۵ سبب ایجاد معنی‌داری در مقدار آماره‌ی F شود. بنابراین براساس نتایج جدول۲، مدل حاصل از روش رگرسیون پیش‌رو از پارامترهای ورودی بجز پارامترهای طوفان تندر، بزرگی باد و سریع‌ترین سرعت باد برای مدل‌سازی طوفان گرد وخاک استفاده می‌نماید.

نتایج آزمون گاما

با استفاده از روش گاماتست برای پیش‌پردازش عوامل، می‌توان ترتیب میزان اهمیت عوامل ورودی و بهترین ترکیب

www.SID.ir

Archive of SID

را از میان تمام ترکیب‌های ممکن برای ایجاد یک مدل هموار برای ورود به شبکه عصبی مصنوعی بدست آورد. در این روش ترکیب تمام عوامل ورودی به نرم‌افزار Wingamma معرفی شده و مقدار گاما تعیین می‌شود. سپس به‌ترتیب هر یک از عوامل از ترکیب حذف می‌شود. در صورتی‌که گاما از گامای بدست آمده در ترکیب تمام عوامل ورودی کمتر باشد، عامل حذف شده در مدل‌سازی اثر منفی دارد و در صورتی‌که بزرگتر باشد، عامل حذف شده در مدل‌سازی اثر مثبت دارد. به هر میزان تغییرات گاما بیشتر باشد اثر مثبت یا منفی عامل بیشتر است. البته به این نکته باید توجه داشت که شاید یک عامل به تنهایی اثر منفی داشته باشد ولی ممکن است در ترکیب با عوامل دیگر اثرش مثبت شود. بعکس این امر نیز صادق است. آزمون گاما یک روش برآورد تراز نویز می‌باشد که از مجموعه داده‌ها بدست می‌آید. آزمون گاما خطای موجود در داده‌ها را با توجه به بهترین مدل حاصل از داده‌ها، برآورد می‌کند.

اشتباه استاندارد (SE): از این عامل نیز می‌توان برای مقایسه کارایی ترکیب‌های مختلف در کنار عامل گاما استفاده کرد. معمولا این مقدار صحت رگرسیون خطی در آماره گاما را بیان می‌کند. اگر این مقدار به صفر نزدیک شود، اطمینان بیشتری از مقدار آماره گاما را نشان می‌دهد. نتایج در جدول ۳ ارائه شده است.

نتایج تحلیل عاملی و تجزیه مؤلفه‌های اصلی

برای استفاده از روش‌های تحلیل عاملی و تجزیه مؤلفه‌های اصلی در گام اول باید داده‌ها استاندارد شود. از آنجا که متغیرهای مختلف در این مطالعه دارای مقیاس‌های اندازه‌گیری متفاوتی می‌باشند، از این‌رو برای بدست آوردن نمره Z و استاندارد نمودن آنها ضروری بود. بدین منظور مقدار میانگین و انحراف متغیرها در هر یک از ماه‌ها و برای هر یک از ایستگاه‌ها تعیین گردیده و بعد استاندارد شده‌اند. کلیه محاسبات با استفاده از نرم‌افزار Minitab انجام شد. در تحلیل عاملی از دو روش حداکثر احتمال (Maximum likhoold) و مؤلفه اصلی (principal component) استفاده شد که در هر ستون بالاترین مقدار به‌عنوان تأثیرگذارترین عوامل انتخاب شد. در این روش منفی بودن عامل در انتخاب عوامل تأثیرگذار نبوده و از منفی عامل چشم‌پوشی می‌شود. در روش تجزیه مؤلفه‌های اصلی از شاخص کوواریانس استفاده شد که نشان‌دهنده تغییرات بین دو متغیر است. در این روش اعداد منفی در نظر گرفته نشد، به این دلیل که نشان‌دهنده تأثیر منفی بر روی متغیر می‌باشد، در این صورت به نتایج مدل جواب قابل قبولی ارائه داده نمی‌شود، به این ترتیب عامل منفی کنار گذاشته می‌شود. نتایج در جدول ۴ و ۵ ارائه شده است.

بعد از انتخاب ترکیب ورودی با هر یک از روش‌های آزمون گاما تست، رگرسیون گام به گام پیش‌رونده، تحلیل عاملی، تجزیه مؤلفه‌های اصلی، ترکیب انتخاب شده به مدل شبکه عصبی وارد شد و با استفاده از دو شاخص ضریب همبستگی و میانگین مربعات خطای مدل مورد ارزیابی قرار گرفته شد. برای تعیین مناسب‌ترین روش برای انتخاب ترکیب ورودی به مدل شبکه عصبی از روش‌های مذکور استفاده شد که در جدول ۶ نتایج روش‌ها به‌طور خلاصه ارائه شده است.

بعد از مقایسه مدل‌ها، با استفاده از مناسب‌ترین مدل (شبکه عصبی) عوامل تأثیرگذار در ماه‌ها و فصل‌های مختلف مورد بررسی قرار گرفت. سپس داده‌ها در هر فصل و ماه‌های مختلف به‌طور جداگانه مرتب شد و عوامل تأثیرگذار در هر ماه و فصل بدست آمد تا مهمترین فصل‌ها و ماه‌ها در برآورد طوفان گرد و غبار انتخاب گردد. فصل تابستان با ضریب همبستگی ۰/۹۰ و مجذور میانگین مربعات خطای ۰/۱۷ و فصل بهار با ضریب همبستگی ۰/۸۸ و مجذور میانگین مربعات خطای ۰/۲۲ شبیه‌سازی با اطمینان بیشتری حاصل می‌شود. همچنین در ماه‌های مختلف مدل شبکه عصبی در ماه اردیبهشت با ضریب همبستگی ۰/۹۷ و مجذور میانگین مربعات خطای ۰/۰۴ و در ماه فروردین با ضریب همبستگی ۰/۹۶ و مجذور میانگین مربعات خطای ۰/۰۰۹ و خرداد، تیر و شهریور با ضریب همبستگی ۰/۹۵ و مجذور میانگین مربعات خطای به‌ترتیب ۰/۰۴، ۰/۱۴ و ۰/۰۴ می‌باشد.

www.SID.ir

Archive of SID

جدول ۱- مقادیر ضریب همبستگی ساده (دو به دو) عوامل هواشناسی

عامل	طوفان تندر	بزرگی باد	پیوستگی باد	توانایی دید	سریع‌ترین سرعت باد	میانگین سرعت	سرعت باد غالب	روز طوفانی
روز طوفانی	۰/۷۵	۰/۷۵	۰/۰۰	۰/۶۹	۰/۷۷	۰/۷۰	۰/۸۶	۱
سرعت باد غالب	۰/۷۲	۰/۶۵	۰/۴۵	۰/۵۸	۰/۵۷	۰/۸۹	۱	
میانگین سرعت	۰/۶۲	۰/۳۵	۰/۱۷	۰/۳۹	۰/۳۲	۱		
سریع‌ترین سرعت باد	۰/۶۱	۰/۹۳	۰/۹۰	۰/۴۹	۱			
توانایی دید	۰/۷۴	۰/۵۹	۰/۵۰	۱				
پیوستگی باد	۰/۴۶	۰/۹۵	۱					
بزرگی باد	۰/۶۱	۱						
طوفان تندر	۱							

جدول ۲- نتایج آزمون رگرسیون پیش‌رونده

عوامل ورودی مدل	P-value	آماره F	ضریب تبیین
سرعت باد غالب	۰/۰۰۰۰۰	۸۷۱۸/۲۹	۷۳/۴۷
سرعت باد غالب، توانایی دید	۰/۰۰۰۰۰	۷۰۶۵/۲۶	۸۰/۳۱
سرعت باد غالب، توانایی دید، پیوستگی باد	۰/۰۰۰۰۰	۶۹۰۹/۸۶	۸۵/۲۷
سرعت باد غالب، توانایی دید، پیوستگی باد، میانگین سرعت	۰/۰۰۰۰۰	۶۸۱۳/۸۰	۸۸/۶۳

جدول۳- نمونه‌ای از نتایج آزمون گاما و انتخاب مناسب‌ترین ترکیب ورودی

ترکیب عوامل	Gamma	SE	V.Ratio
۱۱۱۱۱۱۱	۰/۱۵۹۷۲	۰/۰۱۵۴۰۶	۰/۶۳۸۸۹
۰۱۱۰۱۰۱	۰/۱۴۰۸	۰/۰۱۹۳۹۹	۰/۵۶۳۲
۰۱۱۱۰۰۱	۰/۱۷۰۶	۰/۰۱۲۹۷۸	۰/۶۸۲۴
۰۱۱۰۰۱۱	۰/۱۸۹۲	۰/۰۰۹۰۹۴۸	۰/۷۵۶۸۲
۱۰۱۱۱۰۰	۰/۱۸۸۷	۰/۰۱۲۹۶۱	۰/۷۵۴۸
۱۱۰۰۰۱۱	۰/۱۶۵۵	۰/۰۱۳۱۵۲	۰/۶۶۲۰۱

جدول۴- نتایج آزمون تجزیه مؤلفه‌های اصلی (کوواریانس)

عوامل	PC1	PC2	PC3	PC4	PC5
طوفان تندر	-۰/۰۰۰	۰/۰۰۶	۰/۰۷۴	-۰/۹۹۳	۰/۰۷۷
بزرگی باد	-۰/۰۰۹	-۰/۰۰۰	-۰/۰۰۱	-۰/۰۰۱	۰/۰۰۰
پیوستگی باد	-۰/۰۰۰	-۰/۰۰۷	-۰/۰۱۰	۰/۰۷۴	۰/۹۹۶
توانایی دید	۰/۰۰۹	۰/۰۱۰	۰/۹۹۷	۰/۰۷۵	۰/۰۰۵
سریعترین باد	-۰/۰۰۱	۰/۰۰۱	۰/۰۰۱	-۰/۰۵۱	-۰/۰۴۸
میانگین سرعت	-۰/۰۳۳	۰/۹۹۹	-۰/۰۱۰	۰/۰۰۶	۰/۰۰۶
سرعت باد غالب	۰/۹۹۹	۰/۰۳۳	-۰/۰۱۰	-۰/۰۰۱	۰/۰۰۰

www.SID.ir

Archive of SID

جدول۵- نتایج آزمون تجزیه عاملی (مؤلفه اصلی)

عوامل	Factor1	Factor2	Factor3	Factor4	Factor5
طوفان تندر	۰/۱۵۴	-۰/۰۱۰	-۰/۰۲۰	۰/۰۳۶	۰/۰۰۲
بزرگی باد	-۰/۲۲۴	۰/۶۴۷	۰/۱۹۸	۰/۷۰۱	-۰/۰۲۹
پیوستگی باد	-۰/۲۲۹	-۰/۱۰۱	-۰/۰۱۱	-۰/۰۲۰	-۰/۹۶۸
توانایی دید	۰/۵۱۹	۰/۴۶۱	۰/۵۷۹	-۰/۳۹۸	-۰/۱۵۷
سریعترین باد	۰/۸۰۰	-۰/۲۲۵	-۰/۲۰۲	۰/۵۱۶	-۰/۰۴۸
میانگین سرعت	۰/۱۶۱	۰/۵۷۳	-۰/۷۵۷	-۰/۲۶۷	-۰/۰۴۶
سرعت باد غالب	۰/۰۰۰	۰/۰۲۳	۰/۰۸۳	۰/۰۰۰	۰/۰۰۰

جدول ۴- نتایج روش‌های مختلف به‌منظور تعیین ترکیب ورودی مدل شبکه عصبی

روش تعیین ترکیب ورودی	ترکیب ورودی	R^2	RMSE
تجزیه مؤلفه‌های اصلی (Principal Components Analysis)	پیوستگی باد، سریعترین باد، سرعت باد غالب، دید افقی	۰/۸۳	۰/۰۷۸
رگرسیون پیش‌رونده (Stepwise Regression)	سرعت باد غالب، دید افقی، پیوستگی باد، میانگین سرعت	۰/۸۷	۰/۰۴
تجزیه عاملی (Maximum likelihood)	سریعترین باد، میانگین سرعت، بزرگی باد، پیوستگی باد	۰/۷۳	۰/۱۹۵
تجزیه عاملی (Principal Components)	سریعترین باد، سرعت باد غالب، میانگین سرعت، بزرگی باد، پیوستگی باد	۰/۷۸	۰/۰۹۸
آزمون گاما	بزرگی باد، پیوستگی باد، سریعترین باد، سرعت باد غالب	۰/۸۵	۰/۰۶۵

بحث

انتخاب ترکیب مناسب از عوامل ورودی یکی از مهمترین مراحل ساخت و طراحی هرگونه مدل‌سازی ریاضی و هوشمند است. در این تحقیق از روش رگرسیون پیشرونده، تحلیل عاملی، تجزیه مؤلفه‌های اصلی و گاماتست برای پیش‌پردازشعوامل ورودی و انتخاب ترکیب بهینه از عوامل ورودی برای شبیه‌سازی طوفان گرد و خاک به کمک شبکه عصبی پیشخور پس انتشار با توابع آموزشی لورنبرگ مارکوآت استفاده شد. هر یک از روش‌های مورد استفاده در این مطالعه به‌منظور تعیین ترکیب ورودی، نتایج و متغیرهای متفاوتی را به‌عنوان ورودی معرفی کردند. مناسب‌ترین روش برای تعیین ترکیب ورودی به‌ترتیب رگرسیون پیش‌رونده، متغیرهای سرعت باد غالب، دید افقی، پیوستگی باد و میانگین سرعت باد با R^2=۰/۸۷ و RMSE=۰/۰۴ توشط شبکه عصبی انتخاب شد و آزمون تست گاما را در رتبه دوم متغیرهای بزرگی باد، پیوستگی باد، سریع‌ترین سرعت باد و سرعت باد غالب با R^2=۰/۸۵ و RMSE=۰/۰۶۵ به‌عنوان ورودی شبکه معرفی نمود. مهمترین فصول در شبیه‌سازی پدیده طوفان گرد و غبار در استان یزد فصول بهار و تابستان و ماه‌های اردیبهشت، فروردین، خرداد، تیر، شهریور و مرداد می‌باشد که طوفان گرد و غبار در فصول و ماه‌های مذکور بیشتر اتفاق افتاده و پراکنش داده‌ها در این فصول مناسب‌تر

www.SID.ir

Archive of SID

می‌باشند. مدل شبکه عصبی در فصل تابستان با ضریب همبستگی ۰/۹۰ و مجذور میانگین مربعات خطای ۰/۱۷ و فصل بهار با ضریب همبستگی ۰/۸۸ و مجذور میانگین مربعات خطای ۰/۲۲ شبیه‌سازی با اطمینان بیشتری حاصل می‌شود. همچنین در ماه‌های مختلف مدل شبکه عصبی در اردیبهشت‌ماه با ضریب همبستگی ۰/۹۷ و مجذور میانگین مربعات خطای ۰/۰۴ و در فروردین‌ماه با ضریب همبستگی ۰/۹۶ و مجذور میانگین مربعات خطای ۰/۰۰۹ و خرداد، تیر و شهریور با ضریب همبستگی ۰/۹۵ و مجذور میانگین مربعات خطای به‌ترتیب ۰/۰۴، ۰/۱۴ و ۰/۰۴ می‌باشد. نتایج نشان‌دهنده آن است که هر چه بازه زمانی در مطالعات شبیه‌سازی با استفاده از شبکه عصبی کوچک‌تر باشد، مدل از دقت بالاتری برخوردار می‌باشد. نتایج کلی در این مطالعه حکایت از مناسب بودن شبکه عصبی پیشخور پس انتشار با توابع آموزشی لورنبرگ مارکوآت در صورت تعیین ترکیب ورودی با روش‌های رگرسیون پیشرونده و تست گاما در مواردی که متغیرهای ورودی زیاد باشد، با دقت بالاتری نسبت به روش‌های تجزیه مؤلفه اصلی و تجزیه عاملی می‌باشد. همچنین نتایج این مطالعه با نتایج مطالعات Jamalizade (۲۰۰۸)؛ Huang و همکاران (۲۰۰۶) و Orlovsky و همکاران (۲۰۰۵) که بیان می‌کنند تعداد طوفان در فصل بهار بیشتر می‌باشد، بنظر می‌رسد با شناخت بهتر فرایند این طوفان‌ها بتوان پیش‌بینی‌های دقیق‌تری را با استفاده از این شبکه‌ها انجام داد و می‌توان با درصد بالایی با استفاده از شبکه عصبی، پدیده طوفان گرد و غبار را پیش‌بینی کرد.

منابع مورد استفاده

-خداوردی‌لو، ح.، فتحی، پ. و همایی، م.،۱۳۸۳. تخمین هوشمند منحنی رطوبتی خاک با استفاده از شبکه عصبی مصنوعی. دومین کنفرانس ملی دانشجویی منابع آب وخاک. دانشگاه شیراز، ۲۴-۲۳ اردیبهشت: ۹ ص.

-کیا، م. ۱۳۸۹. شبکه‌های عصبی در MATLAB، چاپ سوم، نشر کیان رایانه سبز، ایران، ۲۲۹ ص.

- Bauer, P., Nouak, S. and Winkler, R., 2007. Fuzzy matematical methods for soil survey and Land evaluation. Journal of soil sciences, 40:477-492.
- Chen, W. N., Dong, Z. B., Li, S. Z. and Yang, T. Z., 1996. Wind tunnel lest of the influence of moisture on the erodibility of loessial sandy loam soil by wind. Journal of Arid Environments, 34: 391-402.
- Caniani D., Pascale, S., Sdao, F. and Sole, A., 2008: Neural networks and landslide susceptibility: a case study of the urban area of Potenza. Natural Hazards, 45:55–72.
- Feng, Q. and Endo Cheng, K. N.;2002. Dust storm in China:A case study of dust storm Variation and dust characteristics. Bulltein of Engineering Geologyand the Environment, 61: 253-261.
- Hsu, K., Gupta, H. V. and Sorooshian, S., 1995. Artificial neural network modeling of the rainfall-runoff process. Water Resources Research, 31(10): 2517-2530.
- Huang, M., Peng, G., Zhang, J. and Zhang, S., 2006. Application of artificial neural networks to the prediction of dust storms in northwest China. Journal of Global and Planetary Change, 52: 216–224.
- Jamalizadeh, M. R., Moghaddamnia, A., Piri, J., Arbabi, V., Homayounifar, M. and Shahryari, A., 2008. Dust storm prediction using ANNs technique (A case study: Zabol city). World Academy of Science: Engineering and Technology, 45: 529-537.
- Lee, S., Ryu, J. H., Lee, M. J., Won, J. S., 2006: The Application of artificial neural networks to landslide usceptibility mapping at Janghung, Korea. Mathematical Geology, 38(2): 199-220.
- Orlovsky, L., Orlovsky, N., and Durdyev, A., 2005. Dust storms in Turkmenistan. Journal of Arid Environments, 60: 83–97.
- Rashki, A., Kaskaoutis, D. G., Rautenbach, C. J., Eriksson, P. G., Qiang, M. and Gupta, P., 2012. Dust storms and their horizontal dust loading in the Sistan region, Iran. Aeolian Research, 5:51–62.
- Song, Z., 2004. A numerical simulation of dust storms in China. Journal of Environmental Modelling and Software, 19:141–151.
- Westwell, I., 1999. Fact finder guide weather. PRC Publishing, 57-58 p.

www.SID.ir

Archive of SID

Determining the best input-combination of artificial neural network in order to identify the wind parameters for predicting dust storms phenomenon (Case Study: Yazd province)

M. Yosefi[1*] and L. Kashi Zenouzi[2]

1*-Corresponding author, M.Sc. Student in Watershed Management, Department of Natural Resources and Desert Studies, Yazd University, Iran, Email: mohsenyosefi67@gmail.com

2- Research Expert, Desert Research Division, Research Institute of Forests and Rangelands, Agricultural Research Education and Extension Organization (AREEO), Tehran, Iran

Received:19/2/2013 Accepted:26/7/2013

Abstract

The aim of this study was to determine some factors affecting dust storms phenomenon using different methods. In order to determine the best-input combination, variable reduction techniques such as factor analysis (maximum likelihood, principal component analysis), Gama test, and multivariate forward regression analysis were used. Each of these methods presented different combinations used by feedforward neural network model, with Levenberg–Marquardt algorithm and multivariate forward regression with R^2=0.87 and RMSE=0.04 was selected as the best suitable combination of neural network model. In addition, monthly and seasonal data were applied by neural network using the best-input combination, and the simulation of dust storm phenomenon was done in summer and spring during the months of April, May, June, July, August and September with a higher correlation coefficient and lower mean square error, due to the good distribution of the dust storm data. The results showed that based on these methods used in this study, dominant wind speed, horizontal visibility, continuity and average of wind speed were the most important factors affecting dust storm phenomenon in Yazd province.

Keywords: Artificial neural network, dust storm, stepwise regression, Gamma test, principal component analysis, factor analysis.

www.SID.ir