

ارزیابی الگوریتم‌های شبکه عصبی، مدل‌های سری زمانی و مدل هیبریدی SARIMA-SETAR در پیش‌بینی سرعت باد ماهانه (مقاله پژوهشی)

- ۱- عباس احمدپور، کارشناس ارشد مهندسی منابع آب، گروه مهندسی آب، دانشکده آب و خاک، دانشگاه زابل، زابل، ایران
 ۲- سیدحسن میرهاشمی، دکتری آبیاری و زهکشی، گروه مهندسی آب، دانشکده آب و خاک، دانشگاه زابل، زابل، ایران
 ۳- مهدی پناهی*، استادیار گروه مهندسی آب، دانشکده کشاورزی، دانشگاه زنجان، زنجان، ایران
 m.panahi@znu.ac.ir

دریافت: ۱۳۹۹/۰۸/۱۴

پذیرش: ۱۳۹۹/۱۲/۱۳

چکیده

در این پژوهش به منظور مدل‌سازی و پیش‌بینی سرعت باد ماهانه شهر دزفول از مدل SARIMA، مدل سری زمانی غیرخطی آستانه (SETAR)، مدل هیبریدی SARIMA-SETAR و همچنین شبکه‌های عصبی مصنوعی استفاده شد. افزون بر این، برای انتخاب متغیرهای ورودی مؤثر در پیش‌بینی سرعت باد با شبکه عصبی از الگوریتم PMI بهره گرفته شد. با استفاده از ضابطه همپل و معیار اطلاعاتی آکائیک، مقدار رطوبت نسبی ماهانه در دو ماه قبل $RH(t-2)$ ، مقدار تبخیر ماهانه در یک ماه قبل $E(t-1)$ ، دمای متوسط ماهانه در سه ماه قبل $Tave(t-3)$ و دمای حداکثر ماهانه در یک ماه قبل $Tmax(t)$ به عنوان متغیرهای ورودی مؤثر برای مدل‌سازی و پیش‌بینی سرعت باد ماهانه مشخص شد. به منظور صحت‌سنجی مدل‌های SARIMA و SETAR و مدل هیبریدی SARIMA*SETAR برآزش یافته از توابع خودهمبستگی، خودهمبستگی جزئی و آزمون استقلال باقی‌مانده‌های مدل (الجونگ-باکس) استفاده گردید. مدل برتر بر اساس حداقل مقدار عددی آماره شوارتز و آکائیک تعیین شدند. همچنین برای مدل‌سازی و پیش‌بینی سرعت باد ماهانه با شبکه عصبی برای لایه ورودی و خروجی تابع خطی و برای لایه مخفی از توابع محرک مختلف با الگوریتم آموزش متفاوت استفاده شد. مدل شبکه عصبی با توپولوژی (۵-۱-۱) با تابع محرک تانژانت سیگموئید و با الگوریتم آموزشی لونیبرگ-مارکوارت در مقایسه با مدل‌های خطی SARIMA و غیر خطی SETAR دارای عملکرد بهتر در پیش‌بینی سرعت باد ماهانه ایستگاه سینوپتیک دزفول شناخته شد. نتایج ارزیابی عملکرد مدل‌ها نشان داد که مدل هیبریدی SARIMA*SETAR(2,2,3) در مقایسه با سایر مدل‌های عملکرد بهتری داشت است. با ضریب تبیین ۰/۹۱ و ریشه میانگین مربعات ۷۲ دارای دقت قابل قبولی در پیش‌بینی سرعت باد ماهانه ایستگاه سینوپتیک دزفول می‌باشد.

واژگان کلیدی: سرعت باد ماهانه، شبکه عصبی، سری زمانی، مدل هیبریدی، الگوریتم PMI.

مقدمه

اثر می‌گذارد و از جمله عوامل مؤثر بر نوسانات اقلیم محسوب می‌شود (شیر غلامی و همکاران، ۱۳۸۳). اطلاع از تغییرات سرعت باد، اهمیت بنیادی در پژوهش‌های مربوط به انرژی باد، احداث ساختمان‌ها و محاسبه نیاز آبی گیاهان دارد [۹]. در ادامه به پژوهش‌های انجام شده در زمینه پیش‌بینی و مدل‌سازی سرعت باد توسط محققین داخلی و خارجی اشاره می‌شود:
 [۱۶] برای پیش‌بینی سرعت باد از مدل‌های ARIMA چندمتغیره و مدل‌های شبکه عصبی برگشتی استفاده

کم آبی یکی از بزرگ‌ترین چالش‌ها در بسیاری از نقاط جهان و از جمله ایران است. کمبود بارش در مناطق مرکزی و عدم نفوذپذیری سریع آب در خاک، منجر می‌شود که بیش از ۷۰ درصد بارش دریافتی تبخیر گردد. یکی از عوامل مهم و مؤثر در افزایش میزان تبخیر، سرعت باد است. به همین دلیل، اندازه‌گیری سرعت باد در مناطق مختلف اهمیت زیادی دارد. عنصر باد از جمله متغیرهای پیچیده اقلیمی است و تغییرات آن در مقیاس‌های کوتاه مدت و بلند مدت بر خصوصیات اقلیم جهانی و منطقه‌ای

الگوریتم‌های شبکه عصبی مصنوعی، مدل‌های سری زمانی خطی^۱ SARIMA و غیرخطی^۲ SETAR و مدل هیبریدی SARIMA-SETAR در پیش‌بینی سرعت باد ماهانه ایستگاه سینوپتیک دزفول است. با توجه به اینکه مدل SARIMA مختص مدل‌سازی داده‌هایی که دوره تناوبی و فصلی از خود نشان می‌دهند می‌باشند و مدل غیرخطی SETAR، توانایی مدل‌سازی پدیده‌های غیرخطی را دارد. افزون بر این، با توجه به اینکه سرعت باد از پدیده‌های هیدرولوژیکی پیچیده‌ای است، بنابراین با به کارگیری مدل هیبریدی مذکور، هم بخش فصلی و هم بخش غیرخطی داده‌های ماهانه را می‌توان همزمان مدل‌سازی و پیش‌بینی نمود.

مواد و روش‌ها

منطقه مورد مطالعه

دزفول شهری است در جنوب غربی ایران با مساحت نزدیک به ۴۷۶۲ کیلومتر مربع، در بخش‌های جلگه‌ای استان خوزستان واقع شده است. با برشمردن بخش‌های روستایی آن دارای گستردگی برابر ۷۸۴۴ کیلومتر مربع است. شهر در ارتفاع ۱۴۳ متری از سطح دریا و از شهرهای شمالی استان خوزستان است. شهر دزفول به دلیل عبور رود دز از آن و پیشینه تاریخی‌اش از اهمیت ویژه‌ای برخوردار است. از لحاظ وسعت دومین شهر بعد از اهواز در استان می‌باشد که از سمت شمال به استان لرستان و از سمت شرق به شهرستان مسجد سلیمان و استان چهارمحال بختیاری و از شمال غربی به شهرستان اندیمشک و از جنوب به شهرستان شوشتر و از سمت غرب به شهرستان شوش محدود می‌گردد.

مدل فصلی خطی SARIMA

مدل SARIMA اولین بار برای سری زمانی فصلی توسط باکس و همکاران (۱۹۹۴) به کار برده شد [۴]. اگر در یک سری زمانی رفتار تناوبی در فاصله‌های زمانی مشخص (S) مشخص شود، این سری زمانی دارای تناوب فصلی است و برای مدل‌سازی آن از SARIMA استفاده

کردند. نتایج نشان داد مدل‌های چند متغیره نسبت به مدل‌های تک‌متغیره و شبکه عصبی برگشتی کارایی بهتری دارند. برای پیش‌بینی سرعت باد از روش شبکه عصبی و روش آماری ARIMA استفاده کردند، نتایج نشان دهنده برتری شبکه عصبی در پیش‌بینی بود [۵]. در تحقیقی به تحلیل داده‌های سرعت باد در طی دوره پنج ساله (۲۰۰۲-۱۹۹۸) در شرق ترکیه پرداخته شد [۳].

در تحقیق دیگری پیش‌بینی سرعت باد با استفاده از یک مدل هیبریدی ترکیبی از مدل شبکه عصبی و مدل تجربی تجزیه می‌باشد، انجام شد. نتایج آن، با مدل شبکه عصبی و ARIMA مقایسه شد، نتایج برتری مدل هیبریدی را نسبت به دو مدل دیگر را نشان داد [۱۳]. افخمی و همکاران (۱۳۹۴) به بررسی امکان پیش‌بینی سرعت باد با استفاده از مدل‌های هیبرید شبکه‌های عصبی و شبکه‌های فازی-عصبی و تئوری موجک در ایستگاه شیراز پرداختند. در آخر کارایی بالای شبکه‌های فازی-عصبی نسبت به سایر مدل‌ها در پیش‌بینی سرعت باد را گزارش دادند [۱]. ارزیابی مدل‌های ARIMA و PARMA برای مدل‌سازی و پیش‌بینی بیشترین سرعت باد در ایستگاه سینوپتیک بندر عباس پرداختند. نتایج پژوهش آنها نشان دهنده برتری مدل PARMA در مدل‌سازی و پیش‌بینی سرعت باد نسبت به مدل اریما بود [۲۱]. روند تغییرات سالانه متغیرهای هواشناسی از جمله سرعت باد در سه ایستگاه شیراز، تبریز و مشهد بررسی شد و مشخص شد که در شیراز روند سرعت باد روزانه کاهش می‌یابد و در مشهد و تبریز روند خاصی وجود ندارد [۱۲]. در تحقیقی، میانگین ماهانه سرعت باد و انرژی آن براساس داده‌های ساعتی ۱۱ ایستگاه سینوپتیک استان اصفهان در دوره اقلیمی ۱۹۹۲-۲۰۰۵ بررسی شد [۱۷]. نتایج نشان داد که سرعت باد در طول ماه‌های سرد کمتر است و با شروع فصل بهار سرعت باد افزایش می‌یابد. نتایج بررسی روند تغییرات سرعت باد در شمال غرب ایران نشان داد که سرعت باد در ایستگاه‌های ارومیه، خوی، تبریز و سنندج روند کاهشی، در ایستگاه‌های اردبیل و زنجان افزایشی و در سقز بدون روند خاصی بوده است [۱۸].

هدف اصلی این پژوهش با توجه به اهمیت مدل‌سازی سرعت باد و کاربرد آن در برنامه‌ریزی منابع آب، عملکرد

^۱- Seasonal Auto-Regressive Integrated Moving Average

^۲- Self-Exciting Threshold Auto-Regressive

معیار اکائیکه اصلاح شده از معیار بی‌زین شوارتز (SBC) استفاده می‌شود.

$$AIC = n \times LN \left(\frac{2\pi RSS}{n} \right) + 1 + 2m \quad (2)$$

$$SBC = n \times LN(MSE) + 2m \times LN(MSE)$$

که در آن: n تعداد کل داده‌ها، $m = (p + q + P + Q)$ ریشه مجموع مربعات RSS و MSE میانگین مربعات باقی‌مانده‌ها است. مدلی که کمترین مقدار AIC و SBC را داشته باشد، انتخاب می‌شود.

تشخیص صحت الگو: برای بررسی صحت برازش مدل-ها، نمودار باقی‌مانده‌ها از نظر نرمال بودن و ایستایی مورد ارزیابی قرار خواهند گرفت. بنابراین، با رسم کردن نمودار ACF باقی‌مانده مدل‌های برازش یافته، چنانچه مقادیر ضرایب خودهمبستگی در باند اطمینان ۹۵٪ قرار بگیرند، مدل برازش یافته کفایت لازم را خواهد داشت.

مدل غیر خطی آستانه (SETAR)

این مدل شامل k بخش $AR(p)$ می‌باشد و مرتبه مدل خودهمبسته در هر رژیم یا قطعه متفاوت است. به طوری که، یک فرآیند بر طبق یک مقدار مشاهده شده (یک آستانه) به فرآیند دیگری تغییر می‌کند. یکی از حالت‌های خاص از مدل TAR مدل $SETAR$ می‌باشد و در ابتدا توسط Tong (۱۹۸۳) ارائه گردید. مدل $SETAR$ که دارای دو رژیم می‌باشد، $SETAR(2;p, \tau)$ ، به صورت زیر ارائه می‌گردد.

$$Y_t = \begin{cases} \alpha_{10} + \sum_{i=1}^p \alpha_{1i} Y_{t-i} + \varepsilon_t & \text{if } Y_{t-d} \leq \tau \\ \alpha_{20} + \sum_{i=1}^p \alpha_{2i} Y_{t-i} + \varepsilon_t & \text{if } Y_{t-d} > \tau \end{cases} \quad (3)$$

این مدل به شکل $SARIMA(p,d,q)(P,D,Q)_s$ می‌شود. مشخص می‌شود که در ساختار آن‌ها (p,d,q) جزء غیر فصلی مدل و (P,D,Q) جزء فصلی مدل و S طول دوره فصل است. با به کارگیری عملگر انتقال به عقب B ، شکل کلی مدل به صورت زیر ارائه می‌شود.

$$\phi_p(B)\phi_p(B^S)\nabla^d\nabla_s^D Z_t = \theta_q(B)\theta_q(B^S)a_t \quad (1)$$

که $\Phi(B^S)$ و $\Theta(B^S)$ به ترتیب چند جمله‌ای بر حسب B^S از درجه P و Q هستند. $\phi(B)$ و $\theta(B)$ به ترتیب چند جمله‌ای‌های مرتبه p و q می‌باشند. ∇_s^D عملگر غیر فصلی و ∇^d عملگر غیر فصلی، d تعداد تفاضل‌گیری غیر فصلی، q مرتبه همبسته میانگین متحرک غیر فصلی، P مرتبه خود همبسته فصلی، D تعداد تفاضل‌گیری فصلی، Q مرتبه میانگین متحرک فصلی و S طول فصل می‌باشد. مدل‌های سری زمانی شامل چهار مرحله است که به صورت تکراری انجام می‌شود و عبارتند از:

شناسایی الگو: در این مرحله با رسم نمودارهای خودهمبستگی (ACF) و خودهمبستگی جزئی ($PACF$) ایستایی در میانگین و واریانس داده‌ها مورد مقایسه قرار داده می‌شوند. تابع خودهمبستگی (ACF) یکی از ابزارهای با اهمیت برای آزمون وابستگی داده‌ها است. رفتار این توابع در نمودار همبستگی نگار به‌عنوان یکی از بااهمیت‌ترین مقیاس‌ها برای تخمین الگوی سری زمانی است. در صورت نایستایی، ابتدا سری مورد نظر با استفاده از سری تفاضلی مناسب و تبدیل داده‌ها از روش باکس-کاکس در میانگین و واریانس پایدار شده، سپس سری ایستا می‌گردد. بنابراین، در این مرحله با آنالیز واریانس داده‌های تفاضلی شده و نیز با استفاده از نمودارهای $PACF$ و ACF مراتب p, q, P و Q مشخص خواهد شد. برازش الگو (برآورد پارامترها) در این مرحله با شناسایی الگوهای مناسب در مرحله قبل، برای مقایسه چند الگو و انتخاب بهترین آنها می‌توان از سنجش اطلاعات اکائیکه^۱ (AIC) استفاده کرد معیار اکائیکه اصلاح شده (AIC) از رابطه زیر بدست می‌آید. علاوه بر

2- Schwartzs Bayesian Criterion

1- Akaikes Information Criterion

صورت جداگانه یک مدل AR با مرتبه p و r برآزش داده می‌شود.

به منظور برآورد یک مدل TAR، ابتدا باید پارامترهای آستانه (τ) و زمان تاخیر (d) برآورد شوند. سپس، داده‌ها به دو رژیم پائین و بالا تقسیم می‌شوند و در آخر، ضرایب معادله‌های AR در هر رژیم بدست می‌آید. با حداکثر کردن تابع لگاریتم-درست‌نمایی مقادیر τ و d از طریق رابطه (۴) محاسبه می‌شود:

$$l(\tau, d) = -\frac{n-p}{2} \{1 + \log(2\pi)\} - \frac{n_1(\tau, d)}{2} \log((\hat{\sigma}_1(\tau, d))^2) - \frac{n_2(\tau, d)}{2} \log((\hat{\sigma}_2(\tau, d))^2) \quad (4)$$

صورتی که مقدار Q از مقدار کای دو جدول کمتر می‌شود، مورد تأیید قرار می‌گرفت.

شبکه‌های عصبی مصنوعی

شبکه عصبی مصنوعی یک مکانیسم محاسباتی است که قادر است با استفاده از اطلاعات یک سری اطلاعات جدید را محاسبه کند. یکی از مرسوم‌ترین شبکه‌های عصبی مورد استفاده در هیدرولوژیکی شبکه عصبی پرسپترون چند لایه است. این شبکه از یک لایه ورودی، یک لایه خروجی و یک یا چند لایه بین آن‌ها که به طور مستقیم به داده‌های ورودی و نتایج خروجی متصل نیستند، تشکیل شده‌اند. واحدهای لایه ورودی تنها وظیفه توزیع میزان‌های ورودی‌ها به لایه بعدی و لایه خروجی نیز پاسخ سیگنال‌های خروجی را مشخص می‌کند. در این دو لایه تعداد نرون‌ها برابر با تعداد ورودی‌ها و خروجی‌ها است و لایه یا لایه مخفی وظیفه ارتباط لایه ورودی به لایه‌های خروجی را برعهده دارند. در شبکه عصبی پرسپترون الگوریتم مشخص برای تعیین تعداد لایه‌های مخفی و تعداد نرون‌ها وجود ندارد و این امر اغلب به روش سعی و خطا انجام می‌شود. یکی از مراحل آماده‌سازی و محاسبات عددی برای تغذیه شبکه‌های عصبی نرمال‌سازی داده‌ها به منظور افزایش حساسیت به بزرگی داده‌ها و افزایش قدرت یادگیری می‌باشد که باعث افزایش توانایی پیش‌بینی خواهد شد. در این پژوهش بردارهای ورودی و

که در آن: ε_t نوفه سفید، p و r به ترتیب مرتبه‌های مدل AR رژیم‌های پائین ($Y_{t-d} \leq \tau$) و بالا ($Y_{t-d} > \tau$)، ضرایب معادله‌ها، τ مقدار آستانه و d زمان تأخیر معادله دو رژیمی بالا را مشخص می‌کند و با حداکثر کردن تابع درست‌نمایی، این مقادیر بدست می‌آید. بنابراین، مقادیر مشاهداتی میانگین دبی روزانه (Y_t) بر طبق معادله ۵، در دو دسته یا رژیم مختلف کمتر یا بیشتر از حد آستانه، قرار می‌گیرند. در نتیجه، برای هر رژیم به

که n تعداد کل داده‌ها، p مرتبه مدل AR، n_1 و n_2 به ترتیب تعداد داده‌های واقع شده در رژیم‌های پائین و بالا، $\hat{\sigma}_1^2$ و $\hat{\sigma}_2^2$ نیز واریانس خطاها حاصل از مجموع مربعات باقیمانده‌ها می‌باشند. پس از آن، برای مقادیر τ و d ثابت شده و با به حداقل رساندن AIC، مرتبه‌های دو مدل پائین و بالا تخمین زده می‌شوند (رابطه ۵).

$$AIC(p, r, \tau, d) = -2l(\tau, d) + 2(p + r + 2) \quad (5)$$

آزمون‌های کفایت مدل‌های خطی و غیرخطی و مدل هیبریدی SARIMA-SETAR

از جمله آزمون مورد استفاده برای بیان کفایت مدل، آزمون الجونگ-باکس است. این آزمون، با محاسبه آماره Q که از توزیع کای دو (χ^2_{k-m}) پیروی می‌کند، به شرح زیر به کار گرفته می‌شود:

$$Q = N(N+2) \sum_{k=1}^L (N-k)^{-1} r_k^2(\varepsilon) \quad (6)$$

طبق این رابطه، N تعداد نمونه، L تعداد تأخیرهای تابع خودهمبستگی و r_k^2 مربع مقدار خودهمبستگی سری‌های زمانی باقیمانده (ε) در تأخیر k ام می‌باشند. آماره Q محاسبه شده با مقدار کای دو مستخرج از جدول در سطح معنی‌داری ۵ درصد مقایسه می‌گردد. کفایت مدل در

برآورد اطلاعات متقابل جزئی (PMI)

با توجه به یک متغیر خروجی تصادفی Y ، مقداری عدم قطعیت در مورد یک مشاهده y که عضوی از Y می‌باشد، وجود دارد که می‌توان آن را با توجه به آنتروپی شانون H ، تعریف کرد [۱۹]. ولی با فرض یک متغیر ورودی تصادفی X که Y وابسته به آن است، مشاهدات متقابل (x,y) این عدم قطعیت را کاهش می‌دهد. اطلاع از x ، اجازه می‌دهد تا مقدار y استنتاج شود و همچنین برعکس. بنا به تعریف اطلاعات متقابل $I(X;Y)$ ، کاهش در عدم قطعیت متغیر Y به دلیل مشاهده X است [۶]. این مسئله به صورت یک قسمت مشترک بین دو دایره نمایش داده می‌شود. این قسمت مشترک جایی است که عدم قطعیت کاهش یافته حول X و Y به ترتیب توسط آنتروپی شرطی $H(X|Y)$ و $H(Y|X)$ مشخص شده است. اطلاعات متقابل (MI) را می‌توان توسط رابطه زیر به صورت مستقیم به دست آورد:

$$I(X;Y) = \iint p(x,y) \log \frac{p(x,y)}{p(x)p(y)} dx dy \quad (8)$$

$p(x)$ و $p(y)$ به ترتیب توابع چگالی احتمال حاشیه‌ای X و Y هستند و $p(x,y)$ تابع چگالی احتمال توام 2 است. در هر صورت به صورت عملی، شکل صحیح توابع چگالی احتمال در رابطه 9 مجهول است. از این رو، از برآورد چگالی‌های احتمال به جای آن استفاده می‌شود. با جایگذاری برآوردهای چگالی احتمال با تقریب عددی انتگرال در رابطه 9 خواهیم داشت:

$$I(X;Y) \approx \frac{1}{n} \sum_{i=1}^n \log \left[\frac{f(x_i, y_i)}{f(x_i) f(y_i)} \right] \quad (9)$$

که در آن: f نمایانگر چگالی برآورد شده بر مبنای نمونه‌ای از n مشاهده از (x,y) است. قابل توجه است که از مبناهای مختلفی برای لگاریتم استفاده می‌شود ولی معمولاً از 2 یا e استفاده می‌شود. اگر مبنای لگاریتم ذکر

خروجی شبکه به کمک رابطه (۷) در بازه (۰ و ۱) استاندارد شدند.

$$y = 0.5 + \left(0.5 \times \frac{x - x_{\min}}{x_{\max} - x_{\min}} \right) \quad (7)$$

که در آن: X داده موردنظر، \bar{X} میانگین داده‌ها، X_{\max} : حداکثر داده‌ها، X_{\min} : حداقل داده‌ها و y : داده نرمال شده می‌باشد. در این پژوهشی از داده‌های ایستگاه سینوپتیک دزفول از سال آماری ۱۳۹۲-۱۳۷۰ استفاده گردید. برای مدل‌سازی و پیش‌بینی پارامتر سرعت باد ماهانه ایستگاه سینوپتیک دزفول از داده‌های ماهانه تبخیر (E); حداکثر دمای مطالب (T_{\max}); میانگین دما (Taverage) حداکثر رطوبت نسبی (RH) و سرعت باد (W) استفاده شد. در آخر با استفاده از الگوریتم PMI متغیرهای ورودی موثر از بین مولفه فوق برای مدل‌سازی و پیش‌بینی سرعت باد ماهانه مورد استفاده قرار گرفتند.

معرفی الگوریتم PMI بر پایه انتخاب ورودی (PMIS)

الگوریتم انتخاب ورودی بر مبنای PMI (PMIS) معرفی شده در این تحقیق، در ابتدا توسط Sharma (2000) برای شناسایی متغیرهای ورودی مؤثر در مدل‌های هیدرولوژیکی بسط داده شد. الگوریتم PMI هر تکرار را با پیدا کردن C_s که مقدار PMI را با توجه به متغیر خروجی بیشینه می‌کند (با توجه به ورودی‌هایی که از قبل انتخاب می‌شوند) و با در نظر گرفتن یک ورودی (C) و یک خروجی (Y) انجام می‌دهد. مفهوم آماری که PMI برای C_s برآورد می‌کند بر مبنای محدوده اطمینان که از توزیع تشکیل شده توسط یک حلقه‌ی خودراه‌انداز^۱، تعیین شده است. اگر ورودی معنی‌دار باشد، C_s به S اضافه شده و انتخاب کردن ادامه می‌یابد، تا جایی که هیچ ورودی معنی‌داری باقی نماند، سپس متعاقباً الگوریتم متوقف خواهد شد.

نمونه، ابتدا یک سری $\varepsilon_y \sim N(0,1)$ ساخته می‌شود و سپس توابع چگالی احتمال حاشیه‌ای f_{ε_y} محاسبه می‌شود. مجموع یکصد هزار تکرار مستقل از سری $\varepsilon_x \sim N(0,1)$ ، مستقل از $\varepsilon_y \sim N(0,1)$ ساخته شده است.

برای هر نمونه از ε_y ، توابع چگالی احتمال حاشیه‌ای f_{ε_x} و f_{ε_y} تخمین زده شدند و متعاقباً $I(\varepsilon_x, \varepsilon_y)$ ارزیابی شد. نتایج مقادیر بحرانی I در جدول ۲ برای سطوح اعتماد مختلف درج شده است. دو ضابطه جایگزین متوقف‌کننده فرمول‌بندی شده که به وسیله آن در هر تکرار، $ICs_{Y,S}$ با مقادیر بحرانی مربوطه $Ib(95)$ و $Ib(99)$ به دست آمده مقایسه شد، که از این دو معیار به جای محاسبات مستقیم از سیستم خودراه‌انداز، برای تعیین اینکه کدام متغیر باید انتخاب یا کدام الگوریتم باید متوقف شود، استفاده خواهد شد. حذف محاسباتی حلقه سیستم خودراه‌انداز باعث می‌شود تا انتخاب متغیرهای ورودی بسیار سریع‌تر انجام گردد.

معیار بر مبنای اطلاعات اکائیکه (AIC)

معیار اطلاعاتی اکائیکه (AIC) عنوان سنجش ارتباط بین دقت فیلتر رگرسیونی و اندازه مجموعه ورودی S ، برای فرمول‌بندی کردن این معیار توقف پذیرفته می‌شود [۲]. مقیاس‌هایی مانند AIC معمولاً به عنوان یک مبنای برای ارزیابی در انتخاب مدل‌ها، مورد استفاده قرار می‌گیرند. رابطه AIC به صورت زیر است:

$$AIC = n \log_{\sigma} \left(\frac{1}{n} \sum_{i=1}^n u_i^2 \right) + 2p \quad (10)$$

که در آن: n تعداد مشاهدات، u_i باقیمانده‌های n و p تعداد متغیرهای مدل است. در حالت رگرسیون خطی، تعداد متغیرها برابر با $K + 1$ است که K برابر با تعداد متغیرها است. برای انتخاب متغیرهای ورودی مؤثر با استفاده از معیار AIC، متغیرهایی انتخاب می‌گردند که حداقل مقدار AIC را داشته باشند.

نشود لگاریتم طبیعی در نظر گرفته خواهد شد. با فرض رابطه ۹ می‌توان در نظر گرفت که برآورد دقیق و مؤثر MI (اطلاعات متقابل) بستگی زیادی به روش بکارگرفته شده در برآورد توابع چگالی احتمال حاشیه‌ای و توام دارد. در مجموع سه معیار توقف الگوریتم PMI وجود دارد که عبارتند از (۱) مقادیر بحرانی جدول‌بندی‌شده، (۲) معیار بر مبنای اطلاعات اکائیکه (AIC) و (۳) معیار آزمون همپل.

مقادیر بحرانی جدول‌بندی‌شده

جدول‌های مقادیر بحرانی ضریب همبستگی (R) به آسانی در دسترس می‌باشند، که برپایه رابطه تحلیلی برای توزیع خطای یک تخمین برای اندازه نمونه فرض شده استوار است. در مورد ضریب همبستگی خطی (R)، توزیع تخمین نمونه از یک توزیع t پیروی می‌کند. جدول‌های مقدارهای بحرانی ضریب همبستگی (R) برپایه توزیع t ، به سادگی ساخته شده‌اند [۷] و مقدار بحرانی برای R برای تعداد نمونه‌ها و یک سطح اعتماد معین را تهیه می‌کنند. با این حال، برخلاف ضریب همبستگی خطی، یک تعریف تحلیلی معادل برای I را نمی‌توان برطبق رابطه ۶ نتیجه‌گیری کرد [۱۰]. از این رو، محققان باید از خودراه‌انداز کردن، برای محاسبه $f(I)$ استفاده کنند [۱۱ و ۹]. توزیع تخمینگر اطلاعات بر مبنای هسته $S^{\bar{}}$ ، برای تعدادی از مدل‌های سری زمانی مورد بررسی قرار گرفت که در این پژوهش، یک جایگزین عملی برای سیستم خودراه‌انداز پیشنهاد شد. به جای استفاده از مقادیر تحلیلی، یک روش برای ساختن جدول‌های مقادیر بحرانی محاسبه شده برای I ، استفاده از شبیه‌سازی می‌باشد. شبیه‌سازی مونت‌کارلو برای تعیین عملی توزیع تخمینگر MI که در گام اول در توسعه یک معیار متوقف‌کننده بر مبنای مقادیر بحرانی تقریبی، استفاده شد [۱۱].

در هر شبیه‌سازی، MI برای یک سری داده‌ها در مقایسه با داده‌های نویز-سفید^۱ گوسین محاسبه خواهد شد (با اندازه نمونه n ، بین ۵۰ تا ۵۰۰۰ مقدار). این محاسبه برای به دست آوردن داده‌هایی با مقادیر بحرانی استفاده می‌شود. این داده‌ها می‌تواند برای آزمون مستقل بودن بر مبنای MI به کار گرفته شود. برای هر اندازه

^۱- White-noise

معیار آزمون همپل^۱ (Z)

روش‌های انتخاب داده پرت، به عنوان یک روش آماری قوی برای تعیین تفاوت قابل توجه یک مقدار X از مقادیر دیگر در مجموعه‌ای از مقادیر X به کار می‌رود. این آزمون انحراف یک مشاهده منفرد را نسبت به میانگین تمامی مشاهدات ارزیابی می‌کند. یک مقدار مشاهده‌ای با داشتن امتیاز Z بزرگتر از ۳ بر اساس قانون 3σ برای توزیع نرمال معمولاً به عنوان داده پرت در نظر گرفته می‌شود. داده‌های پرت مقداری بزرگ‌تر از ۳ برابر انحراف معیار از متوسط مجموعه داده‌ها را دارند.

برای ساختارمندی یک ضابطه متوقف کننده برپایه تشخیص داده‌های پرت برای الگوریتم PMIS، فرض اساسی بر این است که مجموعه داده‌ها دارای نسبتی از متغیرهای اضافی و نامربوط بوده و متغیرهای مهم شناسایی خواهد شد. با این وجود، پنهان‌سازی بالقوه داده‌های پرت مهم است و باید لحاظ شود، با فرض اینکه مجموعه داده‌ها بیشتر از یک متغیر وابسته دارد. به همین علت، یک مقدار Z اصلاح شده، که از فاصله همپل استفاده می‌کند، به جای افزایش کارایی روش به کار گرفته شد. فاصله همپل بر اساس میانه مجموعه ورودی‌ها بنا نهاده شده است. نقطه از کارافتادگی آزمون همپل $n/2$ است [۸] و به عنوان یکی از قدرتمندترین آزمون‌های تشخیص داده پرت با وجود چندین داده پرت شناخته می‌شود [۸ و ۱۴]. آزمون همپل با محاسبه انحراف مطلق از میانه PMI برای همه ورودی‌ها به صورت زیر شروع می‌شود:

$$d_j = \left| I_{C_j Y.S} - I_{C_j Y.S}^{(50)} \right| \quad (11)$$

که در آن: رابطه d_j نمایانگر انحراف مطلق، و $I_{C_j Y.S}^{(50)}$ نشان دهنده میانه PMI برای مجموعه ورودی‌های C می‌باشد. بنابراین فاصله همپل را می‌توان به صورت زیر تعیین کرد:

$$Z_j = \frac{d_j}{1.4826 d_j^{(50)}} \quad (12)$$

که در آن: Z_j نشان دهنده فاصله همپل (Z-score اصلاح‌شده) برای مجموعه ورودی Cj و $d_j^{(50)}$ نشان‌دهنده میانه انحراف مطلق (MAD)، dj است. ضریب $1/4826$ فاصله را به نحوی تغییر داده تا قانون $Z > 3$ را بتوان اعمال کرد، همان طوری که در مورد Z-test متعارف به کار گرفته شده است [۱۴]. با استفاده از این معیار متوقف کننده، الگوریتم PMI بر پایه انتخاب ورودی، دیگر حاوی حلقه خودراه‌انداز نیست، و نیز PMI با هیچ مقدار بحرانی از I مقایسه نمی‌شود. در عوض، مقدار Zs برای برگزیدن Cs تعیین می‌شود و اگر $Zs > 3$ باشد، ورودی انتخاب شده و به S اضافه می‌شود، در غیر این صورت ادامه عملیات الگوریتم انتخاب متغیرهای ورودی متوقف می‌شود.

مدل هیبریدی SARIMA-SETAR

همان گونه که در معرفی روش‌های باکس-جنکینز شرح داده شده است، باقی‌مانده‌های حاصل از برازش مدل‌های خطی دارای رفتار غیر خطی هستند. بنابراین، با بهره‌گیری از مدل غیرخطی برای برازش باقی‌مانده‌های مدل‌های خطی می‌توان به مدل‌سازی و پیش‌بینی بهینه دست یافت. در ادامه مراحل ساخت مدل هیبریدی فصلی و غیرخطی SARIMA-SETAR شرح داده شده است:

- ۱) استفاده از توابع خودهمبستگی و همبستگی برای شناسایی مدل خطی مناسب سری زمانی؛
- ۲) استفاده از توابع خودهمبستگی و همبستگی باقی‌مانده‌ها و آماره عددی پورت مانثو برای شناسایی صحت برازش مدل‌های خطی ساریما؛
- ۳) انتخاب مدل برتر خطی ساریما براساس آماره اکایکه و شوارتز؛
- ۴) استفاده از باقی‌مانده‌های حاصل از برازش مدل‌های خطی به عنوان ورودی به مدل غیرخطی دورژیمی SETAE؛
- ۵- استفاده از آماره پورت مانثو و اکایکه برای شناسایی مدل برتر SETAE.

¹- Hampel distance

بر داده‌های سرعت باد ماهانه ایستگاه سینوپتیک دزفول برآزش داده شد. سپس با استفاده از رفتار این دو تابع، غیرایستایی و فصلی بودن داده‌ها تأیید شد.

در پژوهش حاضر، برای مدل‌سازی و پیش‌بینی سرعت باد ماهانه با مدل SARIMA از نرم‌افزار مینی تب ۱۷ بهره گرفته شد. برای مدل‌سازی سری زمانی SETAR و مدل هیبریدی SARIMA-SETAR از نرم‌افزار R استفاده گردید. با توجه به اینکه داده‌های ماهانه بیشتر رفتار تناوبی یا دوره‌ای از خود نشان می‌دهند، بنابراین از میان مدل‌های مختلف باکس-جنکینز از مدل SARIMA استفاده گردید. یکی از فرض‌های اساسی مدل‌سازی سری زمانی نرمال بودن داده‌ها می‌باشد. بررسی اولیه داده‌ها نشان داد که داده‌ها دارای چولگی به راست و بوده و از توزیع نرمال پیروی نمی‌کنند. بنابراین در پژوهش حاضر از میان توابع تبدیل متفاوت تابع تبدیل لگاریتمی برای نرمال‌سازی استفاده شد. همچنین برای حذف مؤلفه فصلی از تفاضل‌گیری مرتبه ۱۲ و برای ایستایی میانگین از تفاضل‌گیری مرتبه یک استفاده شد.

برای مدل‌سازی و پیش‌بینی سرعت باد ماهانه ایستگاه سینوپتیک دزفول با مدل‌های پیشنهادی شبکه عصبی و مدل خطی SARIMA و مدل غیر خطی SETAR و مدل فصلی و غیرخطی هیبریدی SARIMA-SETAR از داده‌های ماهانه سال‌های ۱۳۸۲-۱۳۷۰ برای مرحله آموزش (واسنجی)؛ و از داده‌های سال‌های ۱۳۸۳-۱۳۹۰ برای مرحله صحت‌سنجی (اعتبارسنجی) و از داده‌های ماهانه سال‌های ۱۳۹۱-۱۳۹۲ برای مرحله آزمون (پیش-بینی) استفاده گردید. هیستوگرام داده‌های اولیه سرعت باد ماهانه ایستگاه سینوپتیک دزفول قبل از استاندارد شدن (نرمال شدن) در شکل ۱ ارائه شده است. همان گونه که مشاهده می‌شود داده‌های اولیه سرعت باد ماهانه ایستگاه سینوپتیک دزفول دارای چولگی به سمت راست است.

معیارهای آماری ارزیابی مدل‌های پیشنهادی

برای ارزیابی و مقایسه دقت مدل‌های پیشنهادی در برآورد سرعت باد ماهانه، از شاخص‌های آماری سنجش خطا استفاده شد. از مهم‌ترین معیارهای آماری سنجش خطا به کار گرفته شده در این پژوهش می‌توان به میانگین خطای مطلق، ضریب تعیین و ریشه میانگین مربعات خطا اشاره کرد که به ترتیب در روابط ۱۳ تا ۱۵ آمده است.

میانگین خطای مطلق

$$MAE = \frac{1}{n} \sum |Q_i - \hat{Q}_i| \quad (13)$$

ریشه میانگین مربعات خط

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (Q_i - \hat{Q}_i)^2}{n}} \quad (14)$$

ضریب تعیین

$$R^2 = \frac{\sum_{i=1}^n (Q_i - \bar{Q})(\hat{Q}_i - \bar{\hat{Q}})}{\sqrt{\sum_{i=1}^n (Q_i - \bar{Q})^2 \sum_{i=1}^n (\hat{Q}_i - \bar{\hat{Q}})^2}} \quad (15)$$

که در آن: n برابر با تعداد داده‌ها، \bar{Q} متوسط سری زمانی سرعت باد ماهانه مشاهده شده Q_i سری زمانی سرعت باد ماهانه مشاهده شده، \hat{Q}_i سری زمانی سرعت باد ماهانه پیش‌بینی شده و $\bar{\hat{Q}}$ متوسط سری زمانی سرعت باد ماهانه پیش‌بینی شده هستند.

نتایج و بحث

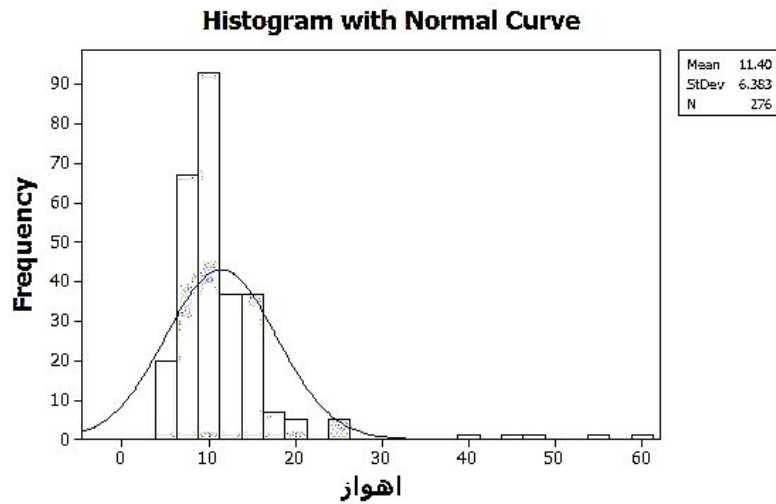
مدل فصلی خطی SARIMA و مدل غیر خطی

SETAR و مدل هیبریدی SARIMA-SETAR

در این پژوهش با استفاده از توابع خودهمبستگی (ACF) ^۱ و خودهمبستگی جزئی $(PACF)$ ^۲ مدل‌های متفاوت فصلی سری زمانی باکس - جنکینز (SARIMA)

1- Autocorrelation Function

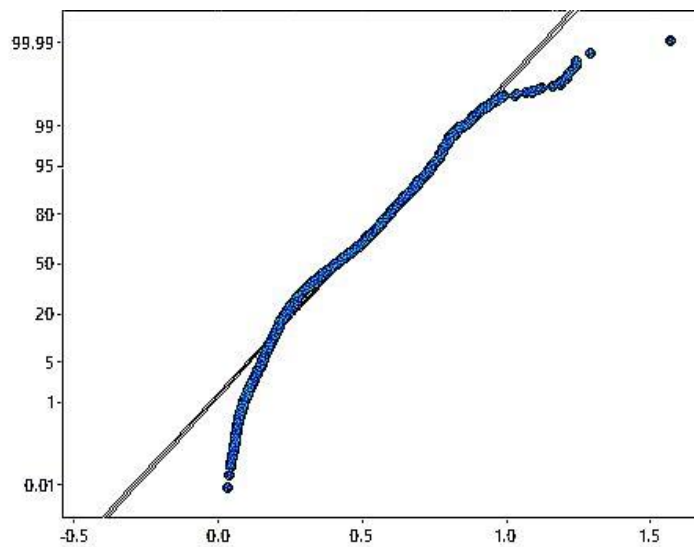
2- Partial Autocorrelation Function



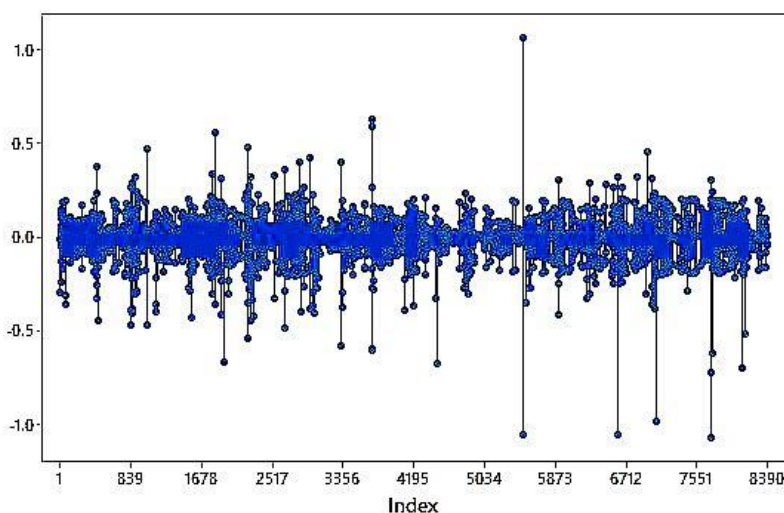
شکل ۱- هیستوگرام داده‌های اولیه سرعت باد ماهانه ایستگاه سینوپتیک دزفول قبل از استاندارد شدن

امتداد خط مستقیم قرار گرفته‌اند، بنابراین، با اطمینان می‌توان گفت که داده‌های مربوطه از توزیع نرمال پیروی می‌کنند. در شکل ۳ همان گونه که مشاهده می‌شود واریانس باقی‌مانده‌ها بدون ساختار خاص است، بنابراین می‌توان از ایستا بودن واریانس مطمئن گردید.

شکل‌های ۲ و ۳ به ترتیب نمودار نرمال شده سری زمانی داده‌های ماهانه سرعت باد بر روی کاغذ احتمال و نمودار سری زمانی داده‌های سرعت باد ماهانه ایستگاه سینوپتیک دزفول پس از ایستا شدن در میانگین و واریانس ارائه شده است. با توجه به شکل ۲، داده‌ها در



شکل ۲- نمودار سری زمانی داده‌های سرعت باد ماهانه ایستگاه سینوپتیک دزفول پس از استاندارد شدن

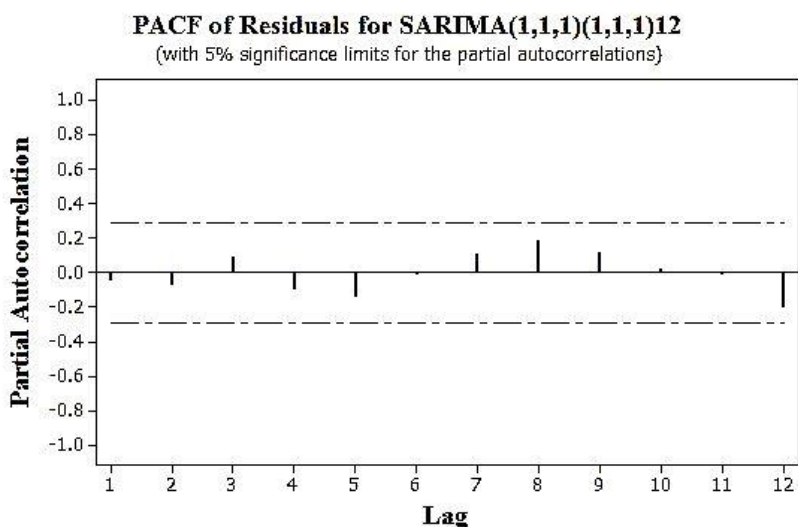


شکل ۳- نمودار سری زمانی داده‌های سرعت باد ماهانه ایستگاه سینوپلتیک دزفول پس از ایستادن

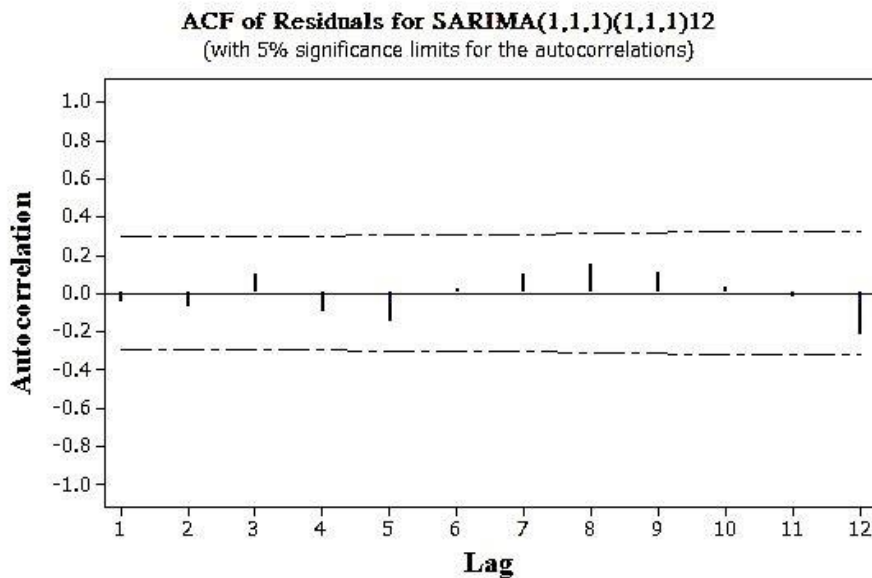
مقادیر ACF و PACF باقیمانده‌ها در محدوده مجاز، باند اطمینان $(\pm 1.96/\sqrt{n})$ ، قرار گرفته است، باقی-مانده‌ها ایستا بودند و هیچ‌گونه روندی را از خود نشان ندادند. جدول ۲، خلاصه‌ای از پارامترهای آماری مدل $SARIMA(1,1,1)^*(1,1,1)_{12}$ برازش یافته سرعت باد ماهانه ایستگاه دزفول ارائه داده است. با توجه به اینکه مقدار آماره پورت مانتو یا الجونگ باکس به دست آمده (Q) کمتر از مقادیر کی دو جدول مشاهداتی (λ^2) است، بنابراین از صحت برازش مدل می‌توان اطمینان نمود.

شکل‌های ۴ و ۵ به ترتیب نمودار تابع خودهمبستگی، خودهمبستگی جزئی باقی‌مانده‌های مدل $SARIMA(1,1,1)(1,1,1)_{12}$ برازش یافته بر داده‌های سرعت باد ماهانه ایستگاه سینوپلتیک دزفول را ارائه داده است. همان‌گونه که ذکر شد یکی از روش‌های آزمودن کفایت مدل برازش شده بر سری زمانی سرعت باد ماهانه، بررسی توابع خودهمبستگی و خودهمبستگی جزئی باقیمانده‌ها است.

شکل‌های ۴ و ۵ به ترتیب نمودار توابع خودهمبستگی و خودهمبستگی جزئی باقی‌مانده‌های مدل SARIMA برازش یافته را نشان می‌دهد. مطابق شکل‌های ۴ و ۵،



شکل ۴- نمودار تابع خودهمبستگی جزئی باقی‌مانده مدل $SARIMA(1,1,1)(1,1,1)_{12}$ برازش یافته بر داده‌های سرعت باد ماهانه



شکل ۵- نمودار تابع خودهمبستگی باقی‌مانده مدل SARIMA(1,1,1)(1,1,1)12 برازش یافته بر داده‌های سرعت باد ماهانه ایستگاه سینوپتیک دزفول

جدول ۲- خلاصه‌ای از پارامترهای آماری مدل SARIMA(1,1,1)*(1,1,1)12 برازش یافته سرعت باد ماهانه ایستگاه دزفول

AIC	SBC	λ^2	Q	θ_2	θ_1	Φ_2	Φ_1	θ_2	θ_1	ϕ_2	ϕ_1
۹۵	۱۲۸	۱۵	۵/۶	۰/۲۸۶۱	۰/۹۳۵	۰/۸۷۹۱	۰/۵۱۰۲	۰/۴۸۱۰	۰/۳۵۹۰	۰/۲۲۵۷	۰/۶۰۷۶

دزفول را تأیید کرد. براساس آماره‌های شوارتز و اکائیکه مدل SETAR(2;6,2) با داشتن حداقل مقادیر آماره‌های مذکور به عنوان مدل سری زمانی غیرخطی SETAR و مدل هیبریدی SARIMA*SETAR(2,2,3) برتر انتخاب گردید. جدول ۳ نتایج آزمون الجونگ-باکس برای آزمودن کفایت مدل‌های غیرخطی SETAR و مدل هیبریدی SARIMA*SETAR برازش یافته ارائه شده است.

مدل SARIMA(1,1,1)*(1,1,1)12 بر اساس حداقل مقادیر آماره شوارتز و اکائیکه به عنوان مدل سری زمانی SARIMA برتر برازش یافته بر داده‌های سرعت باد ماهانه ایستگاه سینوپتیک دزفول تشخیص داده شد. همچنین مقادیر Q آزمون الجونگ-باکس هردو مدل SETAR و مدل هیبریدی SARIMA*SETAR برازش یافته قبول فرض صفر مبنی بر تصادفی بودن سری باقی-مانده‌ها داده‌های سرعت باد ماهانه ایستگاه سینوپتیک

جدول ۳- نتایج آزمون الجونگ-باکس برای آزمودن کفایت مدل‌های SETAR و مدل هیبریدی SARIMA*SETAR برازش یافته

AIC	SBC	نتیجه آزمون	مقدار کای اسکور جدول	Q آماره	تاخیر	تعداد داده	مدل
۱۸۱	۱۵۵	قابل قبول	۶۱/۱۷	۳۴	۴۲	۲۷۶	SETAR(2;1,2)
۲۱۰	۱۶۸	قابل قبول	۶۸	۴۹	۴۲	۲۷۶	SETAR(2,4,1)
۱۷۵	۱۴۷	قابل قبول	۷۹	۵۶	۴۲	۲۷۶	SETAR(2,6,2)
۱۶۸	۱۳۸	قابل قبول	۸۱	۵۵	۴۲	۲۷۶	SARIMA*SETAR(2,2,3)
۱۷۲	۱۲۷	قابل قبول	۸۵	۵۷	۴۲	۲۷۶	SARIMA*SETAR(۲/۳,2)
۱۷۸	۱۳۰	قابل قبول	۹۰	۵۹	۴۲	۲۷۶	SARIMA*SETAR(2,4,1)

مدل شبکه عصبی

به منظور پیش‌بینی متغیر سرعت باد ماهانه در ایستگاه سینوپتیک دزفول با شبکه عصبی مصنوعی، ابتدا اقدام به تعیین متغیرهای مؤثر ورودی برای مدل‌سازی و پیش‌بینی سرعت باد ماهانه گردید. جدول ۴ که شامل متغیرهای ورودی همچون دمای متوسط ماهانه، دمای حداکثر ماهانه، رطوبت نسبی ماهانه و تبخیر ماهانه ایستگاه سینوپتیک دزفول هر یک از متغیرها تا سه ماه قبل در نظر گرفته شدند. بر اساس جدول ۵، با در نظر

گرفتن بیشترین مقدار عددی ضابطه همپل و کمترین مقدار عددی معیار اطلاعاتی اکائیکه در الگوریتم PMI، متغیرهای ورودی مؤثر بر سرعت باد ماهانه در ایستگاه دزفول عبارتند از مقدار رطوبت نسبی ماهانه در دو ماه قبل، مقدار تبخیر ماهانه در یک ماه قبل، دمای متوسط ماهانه در سه ماه قبل و دمای حداکثر ماهانه در یک ماه قبل برای مدل‌سازی و پیش‌بینی سرعت باد ماهانه ایستگاه سینوپتیک دزفول مشخص شدند.

جدول ۴- مشخصات آماری متغیر سرعت باد ماهانه ایستگاه سینوپتیک دزفول

شاخص‌های آماری	تبخیر (E)	حداکثر دمای مطلق (T_{max})	میانگین دما ($T_{average}$)	حداکثر رطوبت نسبی (RH)	سرعت باد (W)
چولگی	۰/۴۵	۰/۱۶	۲/۴۵	۱/۷۸-	۴/۲
میانگین	۲۱۴/۱	۴۰/۳	۲۷	۸۹/۵	۱۰
کشیدگی	-۰/۹۳	۱/۴۱	۹/۳	۳/۳۸	۲۴/۷
بیشینه	۶۲۳/۳	۹۰/۶	۱۰۰	۱۰۰	۵۹
انحراف معیار	۱۵۹/۲۸	۱۰	۱۵/۱	۲۲	۶/۳
واریانس	۲۵۳۷۲/۶	۱۰۰	۲۲۹/۲	۴۸۷/۹	۴۰/۷
کمینه	۱۱/۱	۱۰	۲/۲	۳	۵

جدول ۵- متغیرهای ورودی مؤثر بر متغیر خروجی سرعت باد ماهانه بر اساس الگوریتم PMI

Hampel	AIC(p)	AIC(k)	MC-I*(99)	MC-I*(95)	I(x,y)	Variable	Iteration
۲/۸۳۱۲۲	۱۲/۵۱۹	۲/۴۷۸۳۳-	۰/۰۸۳۸۸۴۵	۰/۰۶۸۳۴۹۱	۰/۰۶۶۴۵۶۵	Log RH(t-2)	1
۲/۶۷۳۴۸	۴۴/۸۸۳۱	۳/۴۱۵۰۲-	۰/۰۹۶۷۵۴۵	۰/۰۶۸۲۷۹۱	۰/۰۷۶۸۲۷۸	Log E(t-1)	2
۱/۹۲۲۳۱	۵۰۵/۲۴۲	۱۸/۰۷	۰/۰۹۴۳۸۱۵	۰/۰۶۸۳۳۹۱	۰/۰۵۴۳۹۵۲	Log Tave(t-3)	3
۱/۳۶۸۰۴	۶۰۶/۵۴	۴۳/۵۱۲۱	۰/۰۷۶۸۸۴۵	۰/۰۶۸۳۸۹۱	۰/۰۵۱۳۸۴۱	Log Tmax (t)	4

به منظور تعیین مقادیر بهینه تعداد نرون‌ها و تعداد لایه‌های مخفی و توابع انتقال در شبکه عصبی مصنوعی ساختارهای متفاوت مورد ارزیابی و بررسی قرار گرفت. تعداد لایه‌ها و همچنین تعداد نرون‌های لایه میانی براساس آزمون سعی و خطا تعیین گردید.

در مدل‌سازی سرعت باد ماهانه با شبکه عصبی مصنوعی سناریوهای مختلف آزمایش شد. به منظور تعیین مقادیر بهینه لایه‌های مخفی و توابع انتقال در شبکه عصبی مصنوعی، مدل‌های متفاوت ارزیابی شد. سپس با توجه به معیارهای ارزیابی مدل، سناریوهای که دارای کمترین خطا و بیشترین همبستگی با داده‌های مشاهده‌ای سرعت باد ماهانه ایستگاه سینوپتیک دزفول به عنوان ساختار بهینه انتخاب شد.

همچنین در این مطالعه برای جلوگیری از آموزش بیش از حد شبکه^۱ که منجر به عدم تعمیم‌پذیری برای داده‌های جدید خواهد شد، داده‌ها به سه سری داده‌های آموزشی، داده‌های صحت‌سنجی و داده‌های آزمایشی تقسیم شدند. از بین توابع انتقال موجود، تابع برای لایه ورودی و خروجی تابع خطی و برای لایه مخفی از بین توابع محرک مختلف همچون تابع تانژانت سیگموئید مناسب تشخیص داده شد. همچنین از الگوریتم‌های آموزشی متفاوت همچون مومنتوم، گرادیان نزولی و لونیگ مارکوات بهره گرفته شد.

بررسی عملکرد مدل‌های شبکه عصبی در پیش‌بینی

^۱ - Over-Fitting

صحت‌سنجی باز هم شبکه مصنوعی با ساختار مذکور ضعیف‌ترین عملکرد را با مقادیر ضریب تعیین و ریشه میانگین مربعات به ترتیب برابر با ۰/۸۴ و ۱۴۳ از بین سایر شبکه‌ها با ساختارهای متفاوت به خود اختصاص داد. بهترین عملکرد در هر دو مرحله واسنجی و صحت‌سنجی مربوط به شبکه عصبی با الگوریتم آموزشی لونیبرگ مارکوات و با ساختار نرونی (۵-۱-۱) بودن در حقیقت مقادیر کم نرون‌ها در لایه میانی و اولی تأثیر بسیار خوبی در بهبود عملکرد شبکه مذکور در مدل‌سازی سرعت باد ماهانه ایستگاه سینوپتیک دزفول داشته است.

سرعت باد ماهانه با توابع آموزش متفاوت و ساختارهای نرونی متفاوت در هر سه لایه (درونی-میانی-بیرونی) در دو مرحله صحت‌سنجی و اعتبارسنجی در جدول ۵ ارائه شده است.

بررسی عملکرد توابع آموزش متفاوت شبکه عصبی نشان داد که ضعیف‌ترین عملکرد در مدل‌سازی و پیش‌بینی سرعت باد ماهانه با شبکه عصبی را الگوریتم آموزشی لونیبرگ مارکوات با ساختار نرونی (۵-۲-۴) با مقادیر ضریب تعیین و ریشه میانگین مربعات در مرحله واسنجی به ترتیب برابر با ۰/۹۱ و ۱۷۲ دارا می‌باشد. در مرحله

جدول ۵- معیارهای ارزیابی عملکرد مدل‌های شبکه عصبی برای پیش‌بینی سرعت باد ماهانه دزفول

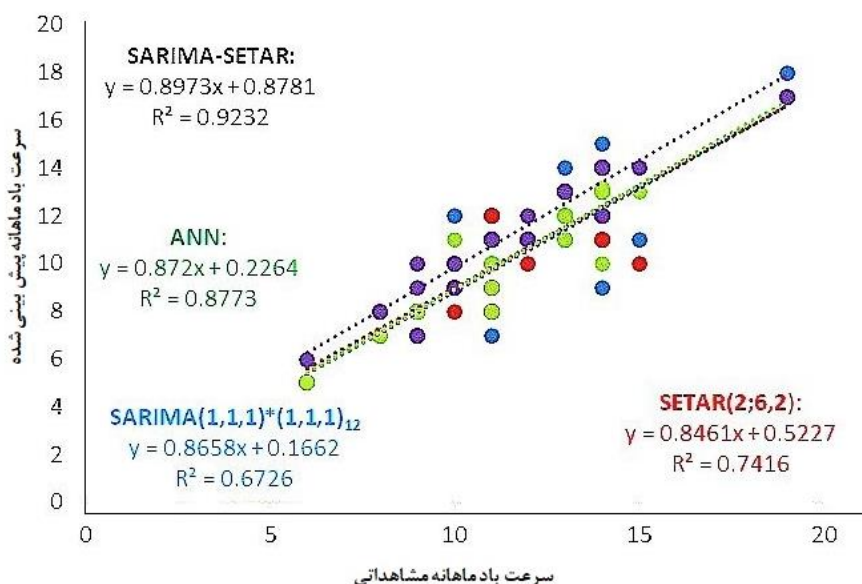
نوع شبکه	نوع آموزش	تعداد نرون لایه اول	تعداد نرون لایه دوم	R ² مرحله آموزش (واسنجی)	RMSE مرحله آموزش (واسنجی)	R ² مرحله صحت‌سنجی (اعتبارسنجی)	RMSE مرحله صحت‌سنجی (اعتبارسنجی)
momentum		۵	۳	۰/۹۲	۱۲۱	۰/۸۷	۹۸
		۳	۴	۰/۹۰	۱۳۴	۰/۸۵	۱۲۰
Conjugate gradient		۷	۲	۰/۹۱	۱۲۸	۰/۸۶	۱۱۷
		۴	۶	۰/۹۳	۱۳۸	۰/۸۸	۱۱۶
		۵	۲	۰/۹۴	۱۴۸	۰/۸۵	۱۲۱
Multilayer Perceptron		۵	۱	۰/۹۵	۱۱۷	۰/۹۲	۱۰۵
		۳	۶	۰/۹۴	۱۴۵	۰/۹۱	۱۰۸
		۵	۴	۰/۹۰	۱۷۲	۰/۸۴	۱۴۳
		۵	۳	۰/۹۱	۱۶۸	۰/۸۸	۱۳۸

ارزیابی عددی این شبکه عصبی در مرحله تست با سایر مدل‌های پیشنهادی در مرحله مذکور مقایسه گردید. علاوه بر این، معیارهای ارزیابی مقایسه مدل‌های پیشنهادی برای پیش‌بینی سرعت باد ماهانه ایستگاه سینوپتیک دزفول در جدول ۶ ارائه شده است. منحنی‌های پراکندگی بین مقادیر مشاهداتی و پیش‌بینی شده متغیر سرعت باد ماهانه دزفول توسط مدل‌های پیشنهادی در شکل ۶ ارائه شده است.

شبکه‌های عصبی با الگوریتم‌های آموزشی گرادیان نزولی و مومنتوم، به ترتیب در رتبه‌های بعدی از نظر عملکرد در مدل‌سازی سرعت باد ماهانه قرار گرفته‌اند. بنابراین، با توجه به برتری نسبی شبکه عصبی با الگوریتم آموزشی لونیبرگ مارکوات و با ساختار نرونی (۵-۱-۱) در هر دو مرحله آموزش (واسنجی) و مرحله صحت‌سنجی (اعتبارسنجی)، ترجیح داده شد که در مرحله پیش‌بینی (تست) از شبکه مذکور استفاده گردد و مقادیر

جدول ۶ - معیارهای ارزیابی مدل‌های پیشنهادی برای پیش‌بینی سرعت باد ماهانه ایستگاه سینوپتیک دزفول

مدل	MAE	RMSE	R ²
SARIMA(۱,۱,۱)*(۱,۱,۱) _{۱۲}	۶۱	۱۰۸	۰/۶۷
SETAR(۲;۶,۲)	۵۳	۹۸	۰/۷۴
ANN	۳۵	۷۸	۰/۸۷
SARIMA-SETAR(۲,۲,۳)	۳۱	۷۵	۰/۹۲



شکل ۶- نمودار پراکندگی مقادیر سرعت باد ماهانه پیش‌بینی شده نسبت به مقادیر مشاهداتی توسط مدل شبکه عصبی مذکور و و مدل خطی $SARIMA(1,1,1)*(1,1,1)_{12}$ و غیرخطی $SETAR(2,6,2)$ و مدل هیبریدی $SARIMA*SETAR(2,2,3)$ بر حسب نات برثانیه

نتیجه‌گیری

سرعت باد یکی از عناصر مهم هواشناسی است. پیش-بینی سرعت باد یکی از عوامل مهم در محاسبه تبخیر، تبخیر و تعرق در یک حوضه آبریز و همچنین برای محاسبه ارتفاع موج ایجاد شده در مخازن سدها در استفاده از توربین‌های بادی از اهمیت زیادی برخوردار است. بنابراین در تحقیق حاضر به منظور پیش‌بینی سرعت باد ماهانه ایستگاه سینوپتیک دزفول از مدل خطی $SARIMA$ ، غیرخطی $SETAR(2;6,2)$ مدل هیبریدی فصلی و غیرخطی $SARIMA-SETAR$ استفاده گردید. نکته قابل توجه در مدل‌سازی سرعت باد ماهانه ایستگاه مذکور با مدل ساریما متفاوت این بود که مدل ساریما با پارامترهای اتورگرسیو و میانگین متحرک در بخش غیرفصلی و فصلی یکسان هستند. که علت این امر شاید وابسته نبودن سرعت باد منطقه مذکور به تغییرات فصلی و پایداری بیشتر شرایط و نوسانات کمتر آب و هوایی می‌باشد.

از سوی دیگر در این پژوهش از الگوریتم PMI برای تعیین مؤلفه‌های ورودی مؤثر در پیش‌بینی سرعت باد ماهانه ایستگاه سینوپتیک دزفول با استفاده از شبکه عصبی مصنوعی بهره گرفته شد. بر اساس آماره‌های همپل و اکائیکه مقدار رطوبت نسبی ماهانه در دو ماه قبل،

مقدار تبخیر ماهانه در یک ماه قبل، دمای متوسط ماهانه در سه ماه قبل و دمای حداکثر ماهانه در یک ماه قبل به عنوان مؤلفه‌های ورودی مؤثر برای مدل‌سازی و پیش‌بینی سرعت باد ماهانه ایستگاه دزفول انتخاب شدند. شبکه‌های عصبی با الگوریتم آموزشی مومنتوم ضعیف‌ترین عملکرد از بین سایر شبکه‌ها داشتند. در آخر، با بررسی معیارهای عددی ارزیابی عملکرد مدل‌ها پیشنهادی برتری مدل شبکه عصبی با توپولوژی (۵-۱-۱) با تابع محرک تانژانت سیگموئید و الگوریتم آموزشی لونیبرگ مارکوات نسبت به سایر ساختارهای شبکه عصبی با الگوریتم‌ها و توابع محرک متفاوت و دو مدل خطی $SARIMA(1,1,1)(1,1,1)$ و غیرخطی $SETAR(2;6,2)$ با مقادیر ضریب تعیین 0.87 و میانگین خطای مطلق ریشه میانگین مربعات 78 در پیش‌بینی سرعت باد ماهانه ایستگاه سینوپتیک دزفول مشخص شد. مدل هیبریدی $SARIMA*SETAR(2,2,3)$ با ضریب تعیین 0.91 و ریشه میانگین مربعات 72 در مقایسه با سایر مدل‌های پیشنهادی عملکرد قابل قبولی داشت. لازم به ذکر است که با افزایش یا کاهش نرون‌ها از ساختار نرونی فوق عملکرد شبکه تغییر محسوس پیدا نکرد.

References

- [1]. Afkhami, H., Talebi, A., Mohammadi, M. Fotouhi F. (2015). Investigation of the feasibility of wind speed prediction using hybrid model of neural networks, neural - fuzzy networks and wavelet (Case study: station of Yazd). *jwmseir*. 9 (30): 31-40. (in Farsi)
- [2]. Akaike, H. (1974). A new look at statistical model Identification. *IEEE. Transaction on Automatic Control AC-19*, 716-723.
- [3]. Akpinar, E.K. and Akpinar, S. (2005). *Energy Conversion and management*. 46, (4), 515-532.
- [4]. Box, G.E.P., Jenkins, G.M. and Reinsel, G.C. (1994). Time series Analysis: Forecasting and Control. 3rdEd. *prentice Hall, Englewood Cliffs Inc.*, New Jersey. 598p.
- [5]. Cadenas, E. and Rivera, W. (2007). Wind Speed forecasting in the south coast of Oaxaca, Mexico. *Renewable Energy*. 32(12)2116-2128.
- [6]. Cover, T.M. and Thomas, J.A. (1991). Elements of Information Theory. In: Wiley Series in Telecommunications. *Wiley & Sons, Inc Journal.*, New York.
- [7]. David, F.N. (1966). Tables of the correlation coefficient. In: Pearson, E.S., Hartley, H.O. (Eds.) *Biometrika Tables for Statisticians*, third ed., vol. 1. *Cambridge University Press, Cambridge*.
- [8]. Davies, L. and Gather, U. (1993). The identification of multiple outliers. *Journal of the American Statistical Association* 88 (423), 782-792.
- [9]. Ghahreman, N. and Gharekhani, A. (2010). Trend analysis of mean wind speed in different climatic regions of Iran. *Iranian Journal of Irrigation and drainage*. No. 1, Vol. 4, p. 31-43. (in Farsi)
- [10]. Goebel, B., Dawy, Z., Hagenauer, J. and Mueller, J.C. (2005). An approximation to the distribution of finite sample size mutual information estimates. In: *IEEE International Conference on Communications (ICC-05), Seoul, South Korea*.
- [11]. Granger, C.W., Maasoumi, E. and Racine, J. (2004). A dependence metric for possibly nonlinear processes. *Journal of Time Series Analysis* 25 (5), 649-669.
- [12]. Khordadi M.J., Eslamian. s.s. and Abedi Kupai. J. (2008). Investigation of the process of meteorological parameters in several regions of Iran. *Technical Workshop on Climate Change Impacts on Water Resources Management*. (in Farsi)
- [13]. Liu, H. Tian, H. and Li, Y. (2012). Comparison of tow new ARIMA-ANN and ARIMA-Kalman hybrid methods for wind speed prediction *Applied Energy*. 98, 415-424.
- [14]. Pearson, R.K. (2002). Outliers in process modeling and identification. *IEEE Transactions on Control Systems Technology* 10 (1), 55-63. doi:10.1109/87.974338.
- [15]. Pearson, R.K. (2002). Outliers in process modeling and identification. *IEEE Transactions on Control Systems Technology* 10 (1), 55-63.
- [16]. Qing, C. Ewing, B. T. and Thompson, M. A. (2012). Forecasting wind speed with recurrent neural networks. *European Journal of Operational Research*. Volume 221(1): 148-154.
- [17]. Rahimzadeh, F., Pedram, M., Sedaghat Kardar, A. and Kamali G.A. (2009). Estimation of wind energy in synoptic stations of Isfahan province. *Journal of Humanities Research*, University of Isfahan. Volume 20, Number 3. P: 156-172. (In Farsi)
- [18]. Rezaei Banafsheh Daragh, M., Javan, Kh. and Zeinali B. (2011). Investigation of wind speed changes in northwestern Iran. *Natural geography*. Volume 4, Number 13. P: 27-36. (in Farsi)
- [19]. Shannon, C.E. (1948). A mathematical theory of communication. *Bell System Technical Journal* 27, 379-423.
- [20]. Sharma, A. (2000). Seasonal to interannual rainfall probabilistic forecasts for improved water supply management: part 1 – a strategy for system predictor identification. *Journal of Hydrology* 239, 232-239.
- [21]. Tamrodi, M.N., Khalili. K., Nazeri Torodi. Z. and Shahnazi. M. 2014. Characteristics of the authors of the article Evaluation of ARIMA and PARMA models for modeling and predicting the maximum wind speed (Case study: Bandar Abbas

Synoptic Station). *National Conference on Applied Research in Science and Engineering*. (In Farsi)

[22]. Tong, H. (1983). *Threshold Models in Non-Linear Time Series Analysis*. Springer, New York.

Evaluation of neural network algorithms, and time-series models and SARIMA-SETAR hybrid model in Monthly wind speed prediction

- 1- Abbas Ahmadpour, Dept. Water Science Engineering, Faculty of Soil and Water, University of Zabol, Zabol, Iran.
- 2- Seyed Hassan Mirhashemi, Dept. of Water Engineering, Faculty of Water and Soil, University of Zabol, Zabol, Iran.
- 3- Mehdi Panahi*, Assistant Professor, Dept. of Water Engineering, Faculty of Agriculture, University of Zanj, Zanj, Iran.

m.panahi@znu.ac.ir

Received: 04 Nov 2020

Accepted: 03 Mar 2021

Abstract

The aim of this study is modeling and predicting the monthly wind speed of Dezful. Therefore, SARIMA model, threshold nonlinear time series model (SETAR), SARIMA-SETAR hybrid model, and also artificial neural networks were used. In addition, the PMI algorithm was used for selecting the effective input variables in predicting wind speed for the neural network model. Using the Hempel criterion and the AIC information criterion monthly relative humidity in the previous two months RH (t-2), monthly in the previous month evaporation (t-1), the average monthly temperature in the previous three months Tave (t-3) and maximum monthly temperature in the previous month Tmax (t) as effective input variables for modeling and predicting monthly wind speed were identified. To validate the SARIMA and SETAR models and the fitted SARIMA-SETAR hybrid model, the functions of autocorrelation, partial autocorrelation and the residual independence test of the model (Ljung-Box) were used. The superiority models were determined based on the minimum numerical value of Schwartz and AIC statistics. For modeling and predicting wind speed with neural network, linear function was used for input and output layer and different stimulus functions with different training algorithms were used for the hidden layer. The neural network model with topology (1-3-5) with sigmoid tangent stimulus function and Levenberg-Marquardt training algorithm was found to have better performance in predicting the monthly wind speed of Dezful synoptic station compared to SARIMA linear and SETAR nonlinear models. The results of the models showed that the SARIMA-SETAR hybrid model (2, 2, 3) had better performance compared to other models. It has acceptable accuracy in predicting the monthly wind speed of the Dezful synoptic, with a coefficient of determination of 0.91 and a root of mean squares 72.

Keywords: Monthly wind speed, Neural network, Time series, Hybrid model. PMI algorithm.