

بررسی الگوی تغییرات ساعتی آلاینده‌ی ازن متناسب با تغییرات دما و

سرعت وزش باد در اصفهان با استفاده از مدل‌های سری زمانی

فاطمه آبروش، کارشناسی ارشد علم داده

دی ماه 1403

## چکیده

تغییرات سری زمانی میزان آلاینده‌ی ازن متناسب با دما و سرعت وزش باد یکی از پارامترهای مهم اقلیمی در بررسی فرآیندهای هیدرولوژیکی، کشاورزی، محیط زیست، بهداشت، صنعت و اقتصاد است. ارزیابی و پیشبینی میزان آلاینده‌ی ازن و دما و سرعت وزش باد کمک شایانی برای مدیران و برنامه‌ریزان هواشناسی، کشاورزی و منابع آب است. یکی از روشهای بررسی داده‌های سری زمانی استفاده از مدل‌های آماری است. با توجه به اهمیت موضوع، در این مقاله میزان آلاینده‌ی ازن متناسب با دما و سرعت وزش باد به صورت ساعتی از 1398/01/01 الی 1402/12/29 در ایستگاه هواشناسی اصفهان با استفاده از مدل‌های SARIMAX، VAR، LSTAM مورد بررسی قرار میگیرد. در ادامه برای بررسی ایستایی مدل از توابع خودهمبستگی (ACF)، خودهمبستگی جزئی (PACF)، آزمون‌های آماری DIKIFULER و ARCH و روش تفاضل‌گیری استفاده شده است. برای بررسی دقت از معیارهای اطلاع بیزی (BIC)، مجذور میانگین مربعات خطا (RMSE) و آکائیک (AIC) استفاده شده است. نتایج بررسی مدل با حافظه بلندمدت سربهای زمانی میزان آلاینده‌ی ازن متناسب با دما و سرعت وزش باد در سطح 95 درصد مورد تأیید قرار گرفت. همچنین تحلیل ساختارهای مختلف نشان‌دهنده آن است که داده‌های میزان آلاینده‌ی ازن متناسب با دما و سرعت وزش باد با استفاده از مدل LSTM بهترین برازش یا عملکرد را دارند که حاکی از دقت مناسب مدل ارائه شده است و می‌توان از آن برای پیش‌بینی در آینده استفاده کرد.

کلید واژه‌ها: آلاینده‌ی ازن، دما، سرعت وزش باد، سربهای زمانی، مدل LSTM، معیارهای آماری.

پیش‌بینی کمیت و کیفیت تحولات اقلیمی یکی از مسائل پیچیده‌ای است که اذهان اقلیم شناسان را به خود مشغول کرده است. اکنون به مدد دستیابی به فناوریهای نوین و در اختیار داشتن سریهای متعدد داده های لازم از متغیرهای اقلیمی و به یاری دانش درک روابط بین این متغیرها، گامهای اساسی در درک و پیش‌بینی روندهای اقلیمی پدیدار شده است. به نحوی که اکنون مدل های کامپیوتری همگی در حیطه توانائی خود به مسائل پیش بینی اقلیم و عوامل مؤثر بر تغییر اقلیم پاسخ می‌دهند. پیش بینی‌های کاملاً دقیق و با خطای صفر، صرف نظر از حوزه و موضوع مورد نظر، بسیار دشوار و تقریباً غیرممکن است، به ویژه آنکه فرایند پیش‌بینی، در محیط کاملاً پیچیده و در ابر غلیظی از عدم قطعیت ها و بازیگران و پیشران های متعدد و مؤثر بر محیط انجام می شود. و داده‌ها و اطلاعات مورد استفاده در پیش‌بینی نیز دارای خصوصیات مبهم و خاکستری می باشند (یاورزاده و همکاران، 1396).

مدل های سری زمانی، مدل های تجربی و ابزاری قدرتمند برای نیکویی برازش و پیش بینی رفتار سیستم های اقلیمی و هیدرولوژیکی از قبیل میزان آلاینده‌ها، دما و سرعت وزش بادهستند (Nail and Momani, 2019). همانطور می دانیم نگرش کلاسیک بر حسب مدل سازی سری های زمانی بنابر کاربرد نگرش باکس جنکیز به ایستایی و غیرایستایی سری های زمانی بستگی دارد. اگر چنین سری هایی خصوصیت حافظه بلندمدت را نشان دهند ارزش پیش‌بینی بر اساس مدل های خود رگرسیون میانگین متحرک ARMA و خود رگرسیون میانگین متحرک انباشته ARIMA معتبر نخواهد بود. در صورت وجود حافظه بلند مدت در سری های زمانی، بین مشاهدات سری در فواصل بسیار زیاد، جدا و دور از هم وابستگی معنی داری وجود خواهد داشت که نمایانگر این موضوع است که مشاهدات مستقل از هم نبوده، و همبستگی بین آنها وجود دارد و مشاهدات گذشته به پیش‌بینی داده‌ها کمک خواهند کرد. با این توصیفات وجود حافظه بلندمدت در پارامترهای جوی شکل ضعیف کارایی را نقض کرده و آنگاه تغییرات تصادفی نبوده و قابل پیش‌بینی خواهد بود. در این راستا ال هاشمی (2014) به ارزیابی و پیش‌بینی بارش ماهانه با استفاده از روش های شبکه عصبی، سری زمانی و رگرسیون چندمتغیره در کشور یمن پرداختند. با توجه به نتایج ارزیابی به ترتیب روش های شبکه عصبی مصنوعی، مدل آریم و رگرسیون بهترین مدل ها جهت برآورد بارش هستند.

در این مقاله با توجه به اهمیت موضوع، مدل شبکه عصبی LSTM جهت برازش و پیش‌بینی میزان آلاینده‌ی ازن و دما و سرعت وزش باد در ایستگاه اصفهان مورد استفاده قرار گرفته است.

## داده‌ها و روش‌شناسی

در این پژوهش با توجه به اهمیت موضوع، داده‌های میزان آلاینده‌ی ازن از ایستگاه هواشناسی 25 آبان اصفهان در استان اصفهان و داده‌های دما و سرعت وزش باد از ایستگاه هواشناسی فرودگاه اصفهان در استان اصفهان جمع‌آوری شده است. داده‌های مورد نیاز در این مقاله که شامل میزان آلاینده‌ی ازن و دما و سرعت وزش باد در تمام ساعات شبانه روز از تاریخ 1398/01/01 ساعت 12 بامداد الی 1402/12/29 ساعت 23 است که از سازمان هواشناسی کشور اخذ گردیده است. در تحلیل داده‌ها و بررسی مدل‌های سری زمانی از مدل LSTM برای مدل‌سازی و شبیه‌سازی سری‌های زمانی آلاینده‌ی ازن، دما و سرعت وزش باد در اصفهان و استفاده شده است، که در بخش بعدی به آن پرداخته می‌شود.

## روش کار در مدل‌های سری زمانی

### 1) پیش‌پردازش داده‌ها

قبل از شروع تحلیل، پیش‌پردازش داده‌ها انجام می‌شود. مراحل پیش‌پردازش شامل موارد زیر است:

بررسی داده‌های گمشده و داده‌های پرت:

داده‌های گمشده: بررسی داده‌ها نشان داد که متغیر آلاینده ازن دارای ۸۵۷۴ مقدار گمشده است، در حالی که متغیرهای دما و سرعت وزش باد هر کدام دارای ۵۱ مقدار گمشده بودند. برای جایگزینی داده‌های گمشده، روش‌های مختلفی اعمال شد، از جمله: جایگزینی با میانگین، جایگزینی با میان، جایگزینی با مد و روش نزدیک‌ترین همسایگی (KNN)

با توجه به ماهیت داده‌ها، روش نزدیک‌ترین همسایگی بهترین نتیجه را ارائه کرد. پس از جایگزین کردن داده‌های گمشده، پاکسازی انجام شد و در داده‌ها داده گمشده وجود ندارد.

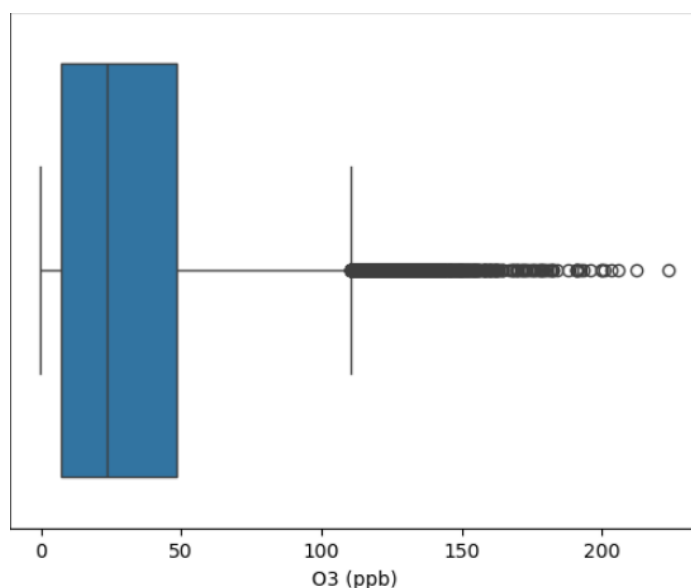
داده‌های پرت: برای شناسایی داده‌های پرت از نمودار جعبه‌ای استفاده شد. اگرچه داده‌های پرت در مجموعه داده مشاهده شد، اما با توجه به ماهیت سری‌های زمانی و وابستگی داده‌ها به زمان، وجود داده‌های پرت قابل قبول در نظر گرفته شد. نمودارهای جعبه‌ای در زیر آورده شده است.

در نمودار آلاینده‌ی ازن جعبه‌آبی نشان‌دهنده دامنه بین چارکی است که شامل 50٪ وسط داده‌ها می‌شود. خط افقی داخل جعبه نشان‌دهنده میانه (Median) یا مقدار مرکزی داده‌ها است. دایره‌هایی که خارج از خطوط قرار دارند، داده‌های پرت هستند و نشان‌دهنده مقادیر بسیار بالا یا بسیار پایین در مقایسه با بقیه داده‌ها هستند. در اینجا داده‌های پرت بیشتری در سمت مقادیر بالا دیده می‌شود. توزیع داده‌ها احتمالاً چوله به راست (مثبت) است

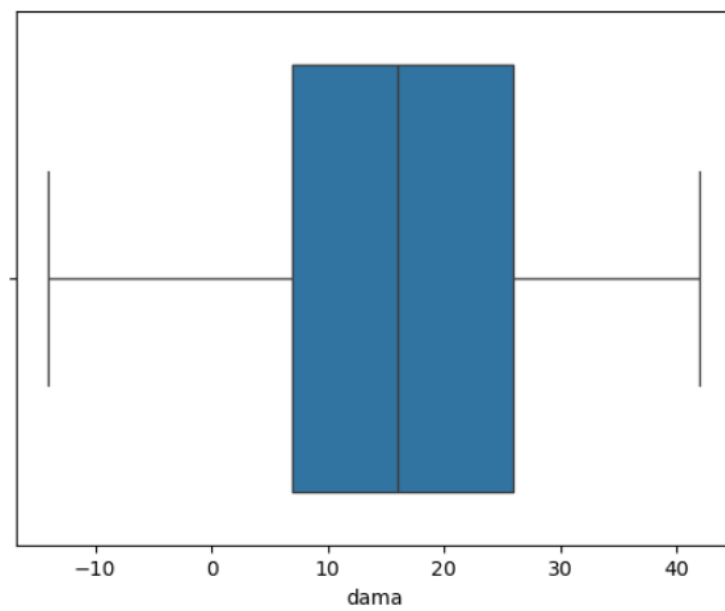
زیرا مقادیر پرت بیشتری در سمت بالا وجود دارد. نمودار هیستوگرام آلاینده‌ی ازن در ادامه رسم شده است که نتایج بدست آمده از نمودار جعبه‌ای را تایید میکند و توزیع متغیر چوله به راست است.

در نمودار متغیر دما داده پرت مشاهده نمی‌شود؛ زیرا هیچ نقطه‌ای بیرون از خطوط وجود ندارد. به نظر می‌رسد توزیع داده‌ها نسبتاً متقارن است زیرا میانه در مرکز باکس قرار دارد و خطوط تقریباً به یک اندازه در هر دو سمت کشیده شده‌اند. این نمودار نشان می‌دهد که متغیر دما یک توزیع نسبتاً متعادل و بدون داده پرت دارد. نمودار هیستوگرام دما در ادامه رسم شده است که نتایج بدست آمده از نمودار جعبه‌ای را تایید میکند و توزیع متغیر نرمال است.

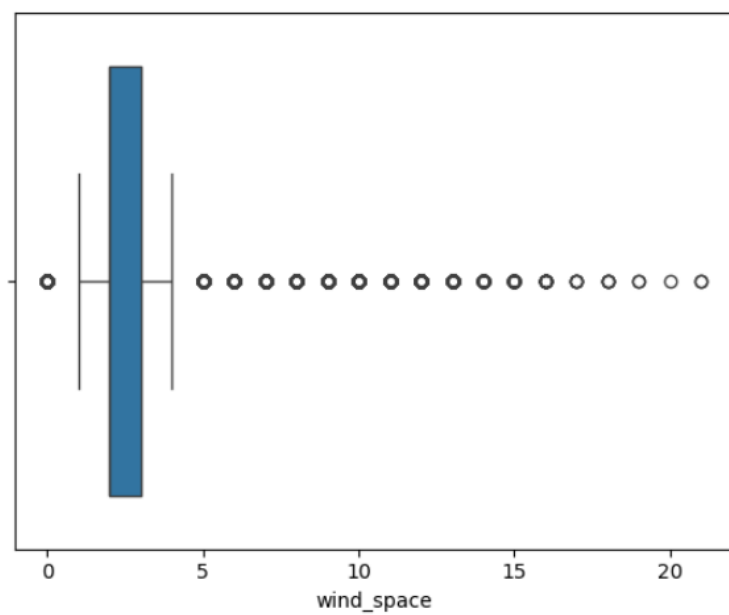
در نمودار سرعت وزش باد دایره‌هایی که خارج از خطوط قرار دارند، داده‌های پرت هستند و نشان‌دهنده مقادیر بسیار بالا یا بسیار پایین در مقایسه با بقیه داده‌ها هستند. در اینجا داده‌های پرت بیشتری در سمت مقادیر بالا دیده می‌شود. توزیع داده‌ها احتمالاً کمی چوله به راست (مثبت) است زیرا مقادیر پرت بیشتری در سمت بالا وجود دارد.



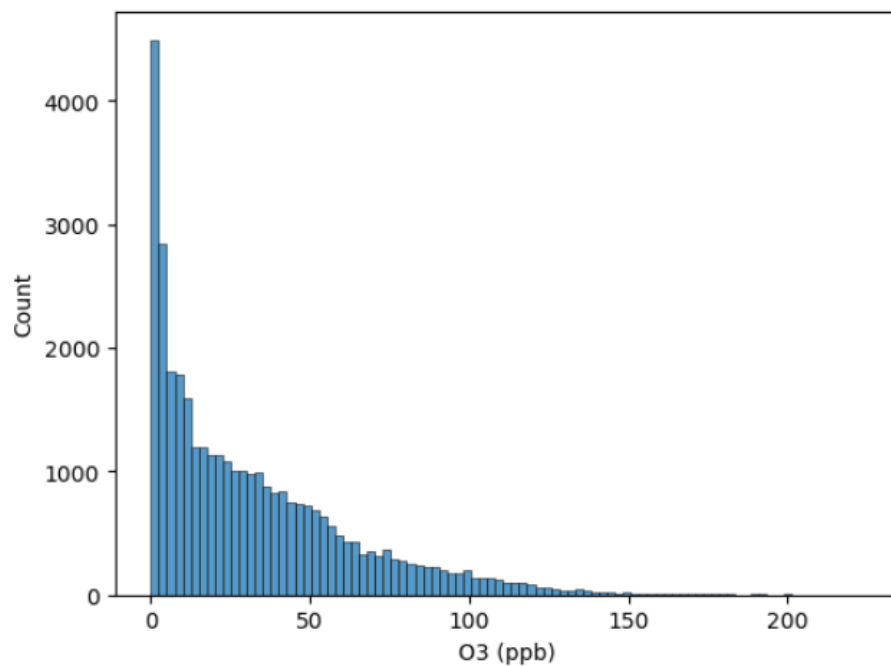
شکل 1- نمودار جعبه‌ای آلاینده‌ی ازن



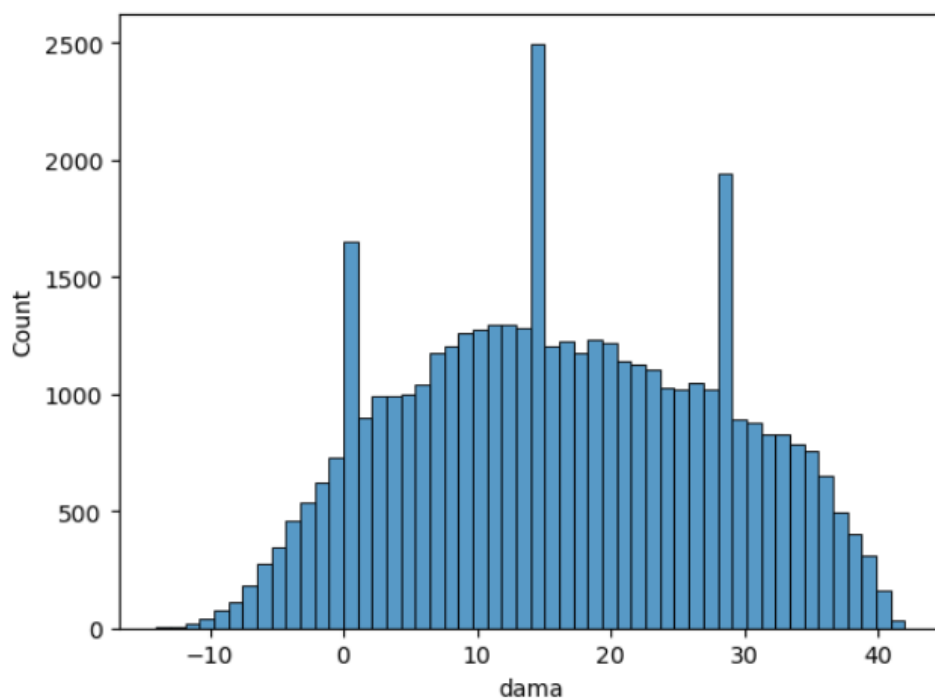
شکل 2- نمودار جعبه‌ای دما



شکل 3- نمودار جعبه‌ای سرعت وزش باد



شکل 4- نمودار هیستوگرام آلاینده‌ی ازن



شکل 5- نمودار هیستوگرام آلاینده‌ی دما

این رویکرد ساختاریافته برای پیش‌پردازش، داده‌ها را تمیز و آماده تحلیل‌های بعدی می‌کند، در حالی که ویژگی‌های ذاتی زمانی داده‌ها حفظ می‌شوند.

تبدیل فرمت تاریخ شمسی به میلادی در ستون مربوط به تاریخ

بررسی سطرهای تکراری

2) بررسی توصیفی داده‌ها

برای تحلیل توصیفی، شاخص‌هایی مانند میانگین، انحراف معیار، کمترین، بیشترین و سایر معیارهای آماری در جدول زیر ارائه شده‌اند:

آلاینده ازن: میانگین مقدار این آلاینده ۳۲.۲۹ است.

دما: میانگین دما برابر با ۱۶ درجه سانتی‌گراد است.

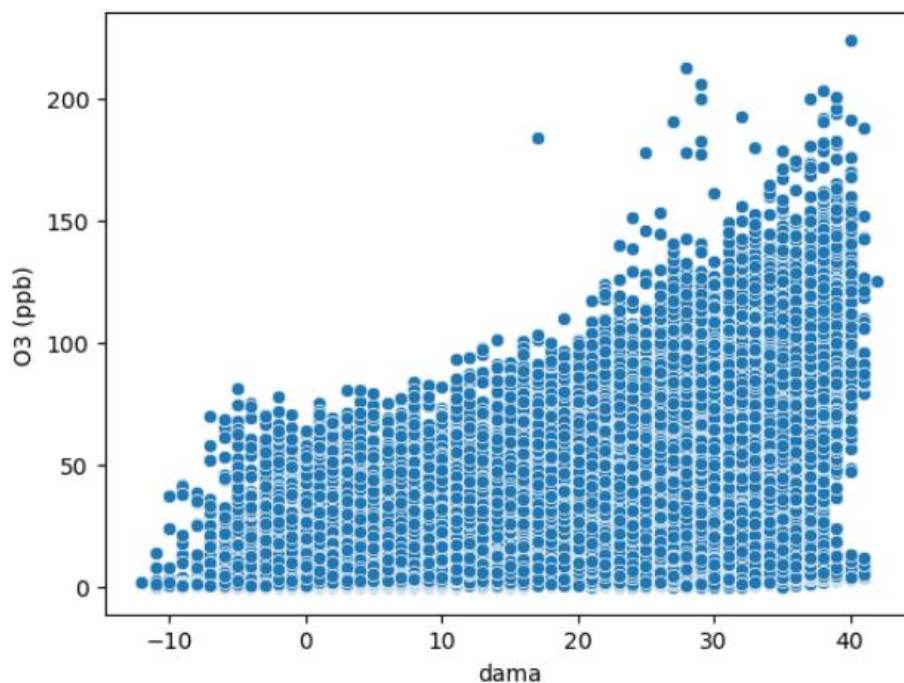
سرعت وزش باد: میانگین سرعت وزش باد ۲.۸۳ متر بر ثانیه است.

	تعداد	میانگین	انحراف استاندارد	کمترین	چارک اول	چارک سوم	بیشترین
آلاینده‌ی ازن	35250.0	32.298115	31.016176	-0.05	7.04	48.42	223.79
دما	43773.0	16.317273	11.637559	-14.00	7.00	26.00	42.00
سرعت وزش باد	43773.0	2.836589	2.260042	0.00	2.00	3.00	21.00

جدول 1- شاخص‌های توصیفی داده‌ها

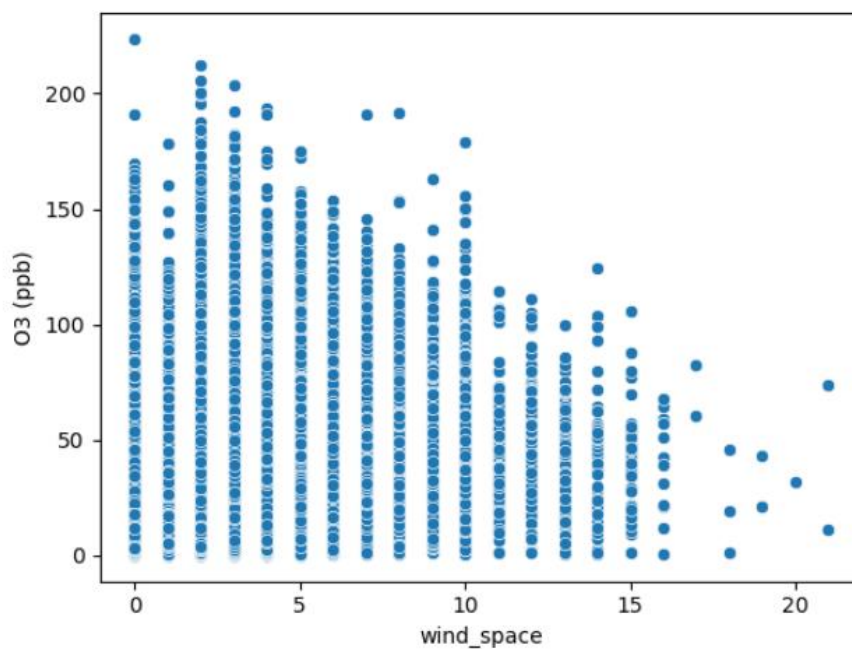
به منظور بررسی همبستگی بین متغیر پاسخ (آلاینده‌ی ازن) و متغیرهای مستقل (دما و سرعت وزش باد) نمودار پراکنش، نمودار حرارتی و آزمون ضریب همبستگی پیرسن انجام شده است که نتایج به شرح زیر است:





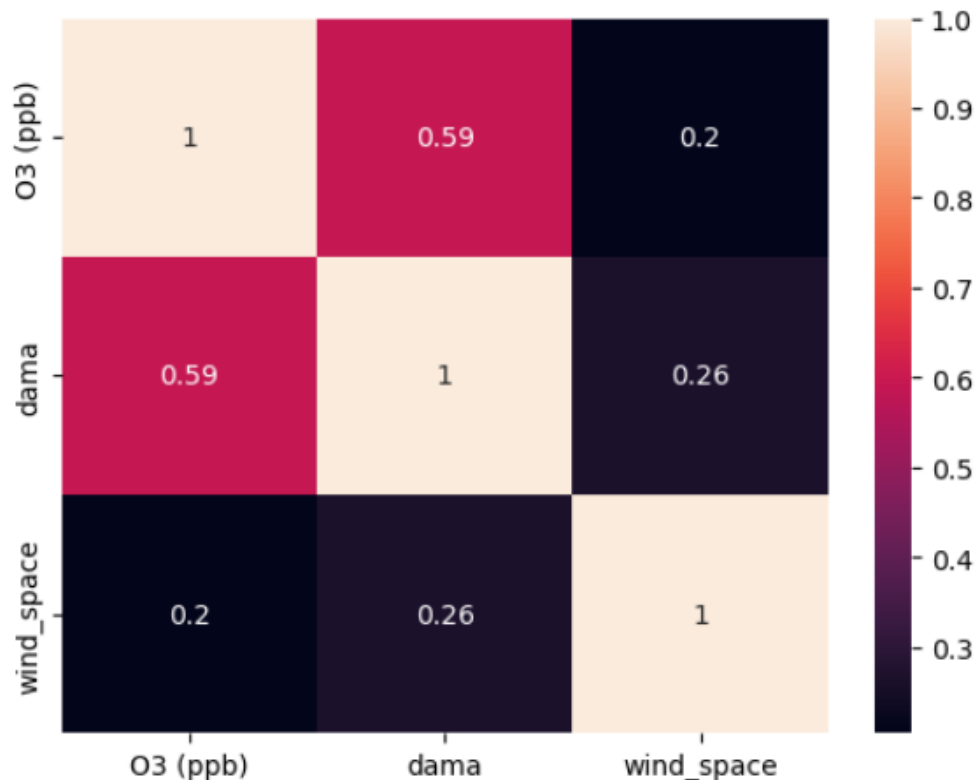
شکل 6- نمودار پراکنش آلاینده‌ی ازن و دما

طبق نمودار پراکنش آلاینده‌ی ازن و دما با افزایش مقادیر دما، مقادیر آلاینده‌ی ازن نیز افزایش می‌یابد. این نمودار پراکندگی نشان می‌دهد که بین این دو متغیر یک رابطه مستقیم وجود دارد.



شکل 7- نمودار پراکنش آلاینده‌ی ازن و سرعت وزش باد

نمودار پراکندگی آلاینده‌ی ازن و سرعت وزش باد نشان می‌دهد که با افزایش مقادیر سرعت وزش باد، مقادیر آلاینده‌ی ازن به طور کلی کاهش می‌یابند و این رابطه معکوس است.



شکل 8- نمودار حرارتی متغیرها

با توجه به نمودار حرارتی بین آلاینده ازن و دما همبستگی قوی با مقدار ۰.۵۹ وجود دارد. این نشان می‌دهد که با افزایش دما، غلظت آلاینده ازن نیز تمایل به افزایش دارد. این رابطه مستقیم می‌تواند ناشی از تأثیر دما بر واکنش‌های شیمیایی تولید ازن در جو باشد.

بین آلاینده ازن و سرعت وزش باد همبستگی ضعیف‌تری با مقدار ۰.۲ مشاهده می‌شود. این نشان می‌دهد که سرعت وزش باد تأثیر کمتری بر غلظت ازن دارد. ممکن است وزش باد به پراکندگی آلاینده‌ها کمک کند، اما تأثیر آن بر ازن به اندازه دما قوی نیست.

بین دما و سرعت وزش باد همبستگی بسیار ضعیفی با مقدار ۰.۲۶ وجود دارد. این نشان می‌دهد که رابطه مستقیم کمی بین دما و سرعت وزش باد وجود دارد. ممکن است در برخی شرایط جوی، افزایش دما با کاهش سرعت باد همراه باشد، اما این رابطه چندان قوی نیست.

آزمون همبستگی پیرسن:

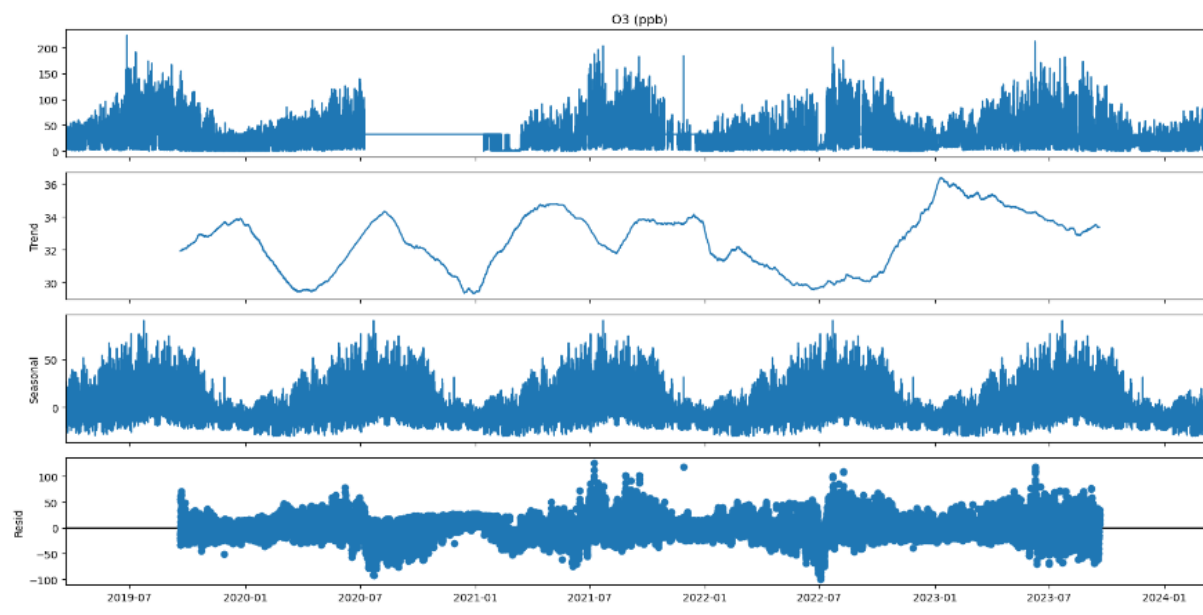
با توجه به مقدار پی-ولیو فرض صفر رد میشود این بدان معناست که بین هر سه متغیر (آلاینده ازن، دما، و سرعت وزش باد) همبستگی معنادار وجود دارد ولی شدت این همبستگی‌ها متفاوت است به طوریکه دما بیشترین تأثیر را بر غلظت آلاینده ازن دارد و سرعت وزش باد تأثیر کمتری دارد، اما همچنان رابطه‌ای معنادار با ازن و دما نشان می‌دهد.

سطح معناداری	ضریب همبستگی پیرسن	
0.0	0.519823	آلاینده‌ی ازن و دما
0.0	0.189258	آلاینده‌ی ازن و سرعت وزش باد
0.0	0.261808	دما و سرعت وزش باد

جدول 2- جدول همبستگی بین متغیرها

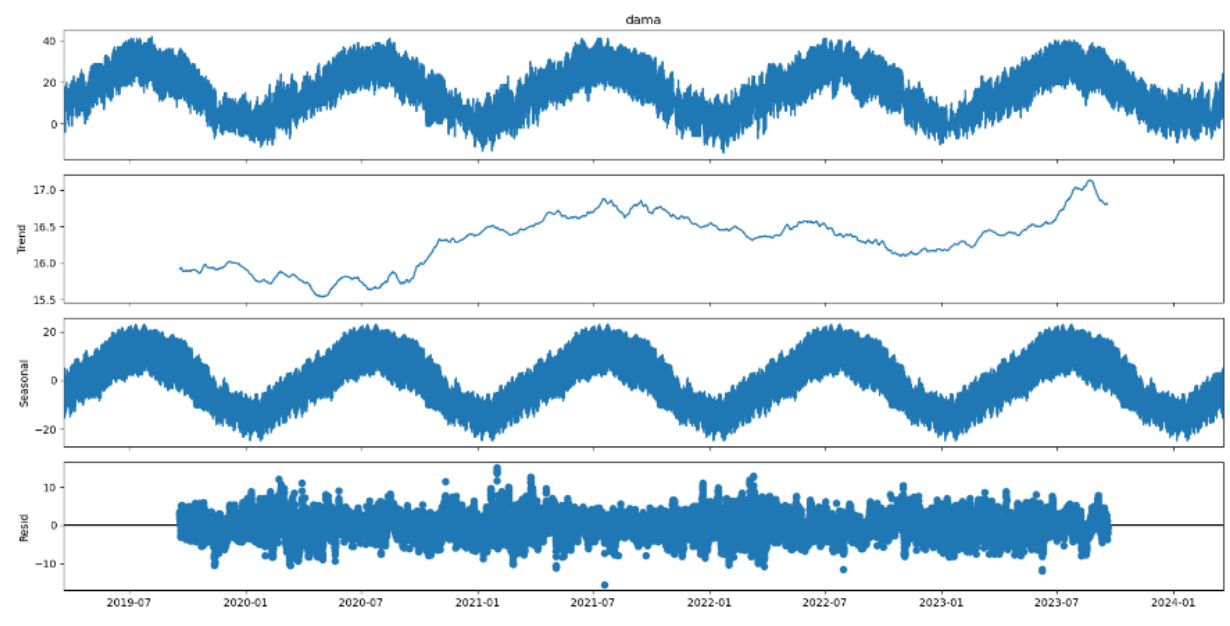
### 3) بررسی استنباطی داده‌ها

در تجزیه و تحلیل سری‌های زمانی، مانایی در میانگین و واریانس یکی از مهم‌ترین شروط برای استفاده از مدل‌های مرسوم است. برای بررسی مانایی سری‌های زمانی آلاینده‌ی ازن، دما و سرعت وزش باد، نمودارهای تفکیکی بر اساس روند، تغییرات فصلی و تغییرات غیرقابل پیش‌بینی رسم شده است. همچنین، به منظور بررسی دقیق‌تر، از آزمون‌های آماری ADF (Augmented Dickey-Fuller) و KPSS (Kwiatkowski-Phillips-Schmidt-Shin) برای بررسی مانایی در میانگین و از آزمون ARCH (Autoregressive Conditional Heteroskedasticity) برای بررسی مانایی در واریانس استفاده شده است. از آزمون هم‌انباشتگی برای بررسی رابطه بلندمدت معنادار بین متغیرها استفاده شده است. نتایج به شرح زیر است:



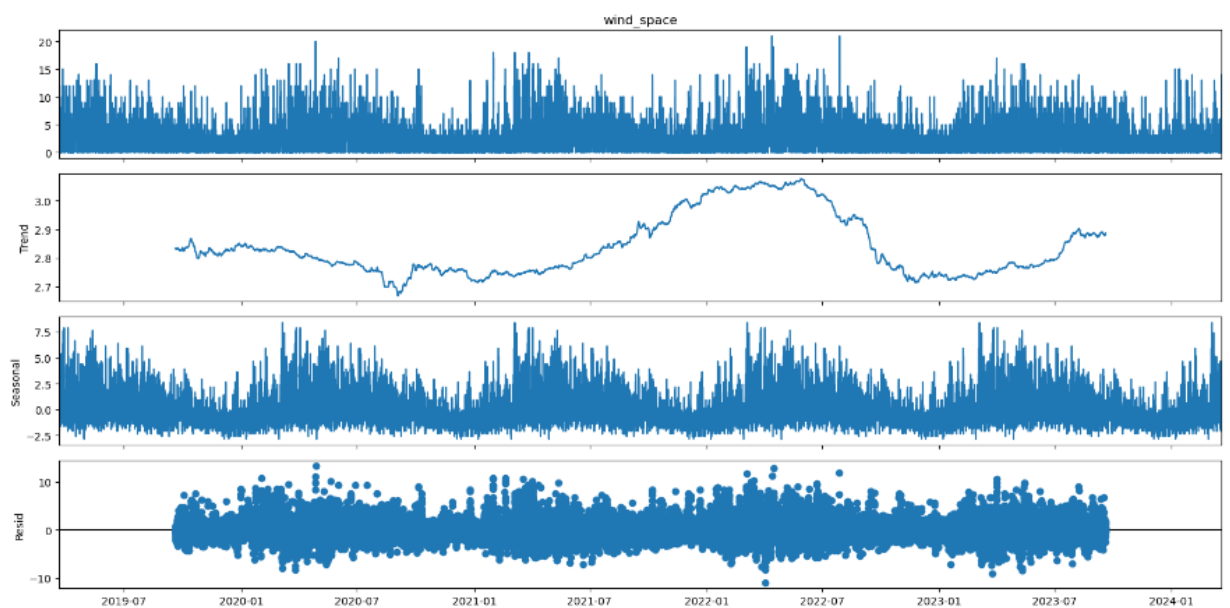
شکل 9- نمودار تفکیک سری زمانی آلایندهی ازن

در نمودار آلایندهی ازن بخش اول نمودار سری زمانی داده‌ها از سال 2019 الی 2024 است. نمودار دوم نمودار روند (Trend) است که روند افزایشی یا کاهشی ندارد و تقریباً روند ثابتی دارد. نمودار سوم تغییرات فصلی (Seasonal) را نشان می‌دهد که نشان‌دهنده الگوهای تکراری در داده‌ها است که در بازه‌های زمانی مشخص (مانند روزانه، ماهانه یا سالانه) رخ می‌دهد. نمودار آخر باقیمانده‌ها (Residual) است که نشان‌دهنده تغییرات غیرقابل پیش‌بینی یا نویز در داده‌ها است که توسط روند و تغییرات فصلی توضیح داده نمی‌شود. باقیمانده‌ها الگوی خاصی ندارند و حول صفر پراکنده‌اند که نشان‌دهنده این است که مدل تجزیه به خوبی داده‌ها را توصیف کرده است.



شکل 10- نمودار تفکیک سری زمانی آلاینده‌ی دما

در نمودار دما به طور کلی روند افزایشی مشاهده می‌شود که نشان‌دهنده‌ی افزایش مقدار متغیر در طول زمان است. همچنین تغییرات فصلی مشهود است. باقیمانده‌ها الگوی خاصی ندارند که نشان‌دهنده‌ی این است که مدل تجزیه به خوبی داده‌ها را توصیف کرده است.



شکل 11- نمودار تفکیک سری زمانی آلاینده‌ی سرعت وزش باد

در نمودار سرعت وزش باد در بازه‌ی زمانی حدوداً 2022/05 نمودار افزایش داشته است و در بازه‌ی زمانی دیگر تقریباً روند ثابتی داشته است. همچنین تغییرات فصلی مشهود است. باقیمانده‌ها الگوی خاصی ندارند که نشان‌دهنده‌ی این است که مدل تجزیه به خوبی داده‌ها را توصیف کرده است.

بررسی ایستایی در میانگین با استفاده از آزمون ADF و KPSS:

طبق نتایج بدست آمده از سطح معنی داری آزمون ADF و KPSS که کمتر از 0.05 هستند فرض صفر (سری زمانی دارای ریشه واحد است یعنی سری زمانی غیر ایستا است) رد می‌شود و سری زمانی در میانگین ایستا است.

متغیرها	آماره‌ی آزمون ADF	سطح معنی داری ADF	آماره‌ی آزمون KPSS	سطح معنی داری KPSS
آلاینده‌ی ازن	-11.248	1.734891164589603e-20	0.758	0.010
دما	-3.854	0.002	0.699	0.013
سرعت وزش باد	-19.139	0.000	0.449	0.000

جدول 3- بررسی مانایی در میانگین

بررسی ایستایی در واریانس با استفاده از آزمون ARCH:

طبق نتایج بدست آمده از سطح معنی داری آزمون ARCH که کمتر از 0.05 هستند فرض صفر (سری زمانی مانا در واریانس نیست) رد می‌شود و سری زمانی در واریانس ایستا است.

متغیرها	سطح معنی داری آزمون ARCH
آلاینده‌ی ازن	0.0
دما	0.0
سرعت وزش باد	0.0

جدول 4- بررسی مانایی در واریانس

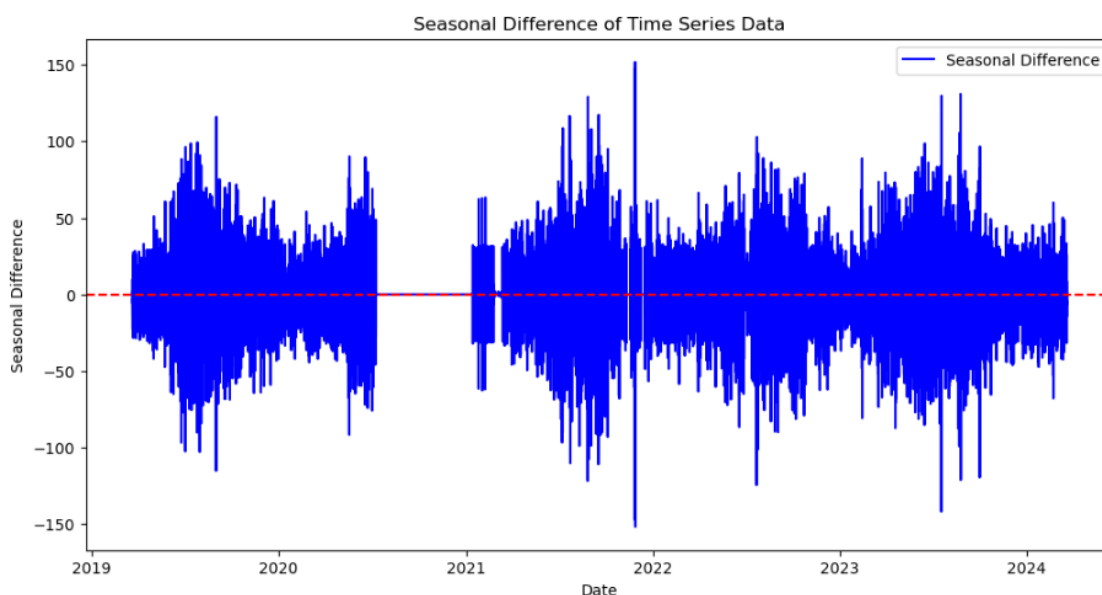
بررسی رابطه بلندمدت معنادار بین متغیرها (آزمون هم‌انباشتگی):

طبق نتایج بدست آمده از سطح معنی داری آزمون Cointegration، فرض صفر رد می‌شود. این موضوع نشان‌دهنده‌ی وجود یک رابطه بلندمدت بین دو متغیر است. هم‌انباشتگی به بررسی ترکیب خطی متغیرها و رابطه‌ی بین آن‌ها می‌پردازد.

متغیرها	سطح معنی داری
آلاینده‌ی ازن و دما	9.820652953759152e-09
آلاینده‌ی ازن و سرعت وزش باد	0.0
دما سرعت وزش باد	0.0

جدول 5- بررسی هم انباشتگی

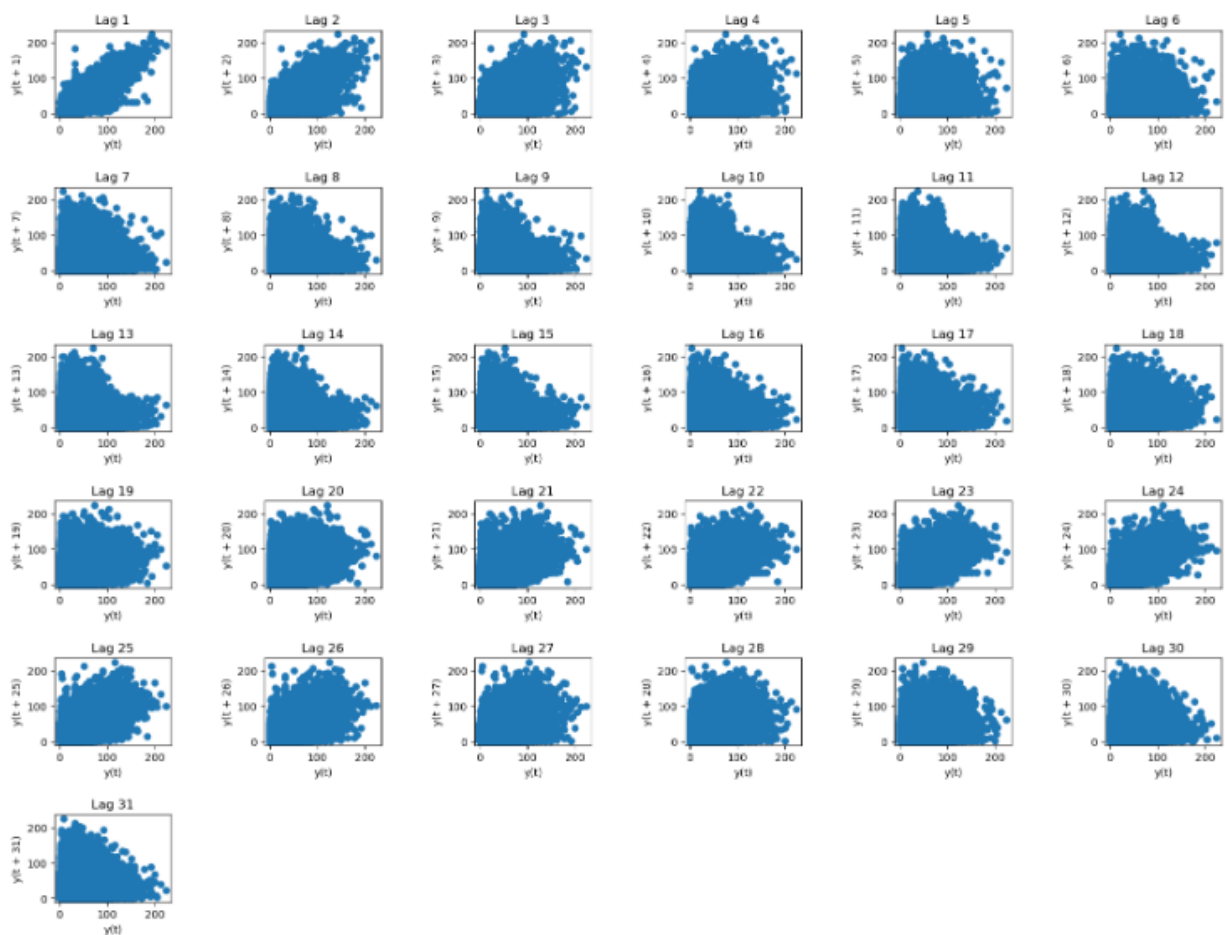
متغیر پاسخ در این مقاله آلاینده‌ی ازن و متغیرهای مستقل دما و سرعت وزش باد هستند. به منظور حذف اثرات فصلی از داده‌ها و بررسی تغییرات غیرفصلی تفاضل گیری مرتبه اول بر روی متغیر مستقل انجام شده است و اثر تغییرات فصلی حذف شده است. مقادیر مثبت نشان‌دهنده‌ی این است که مقدار متغیر در این دوره‌ی فصلی نسبت به دوره‌ی فصلی قبلی افزایش یافته است و مقادیر منفی نشان‌دهنده‌ی این است که مقدار متغیر در این دوره‌ی فصلی نسبت به دوره‌ی فصلی قبلی کاهش یافته است.



شکل 12- نمودار حذف اثر فصلی متغیر پاسخ

بررسی خود همبستگی در متغیر پاسخ:

در نمودارهای زیر خود همبستگی متغیر آلاینده‌ی ازن به ازای تاخیرهای 1 تا 31 بررسی شده است. نتایج حاکی از آن است که متغیر پاسخ با زمان‌های قبلی خود همبستگی دارد پس استفاده از مدل‌های سری زمانی مناسب است.



شکل 13- نمودار همبستگی متغیر پاسخ در تاخیرهای متفاوت

4) مدل سازی:

با توجه به بررسی‌های انجام شده به نظر می‌رسد مدل SARIMAX (Seasonal Autoregressive Integrated Moving Average with eXogenous factors) برای برازش به داده‌ها می‌تواند مناسب باشد. سپس با استفاده از تکنیک‌های مانند تحلیل خودهمبستگی (ACF) و خودهمبستگی جزئی (PACF)، پارامترهای مدل SARIMAX را تعیین می‌کنیم و در نهایت داده‌ها را به دو بخش آموزش و آزمایش تقسیم می‌کنیم. مدل را روی داده‌های آموزش برازش می‌کنیم و روی داده‌های تست عملکرد آن را ارزیابی می‌کنیم. پارامترهای مدل شامل:

- $p$ : درجه مدل خودرگرسیون (AR)

- $d$ : درجه تفاضی گیری (I)

- $q$ : درجه مدل میانگین متحرک (MA)



- P: درجه مدل خودرگرسیون فصلی (SAR)

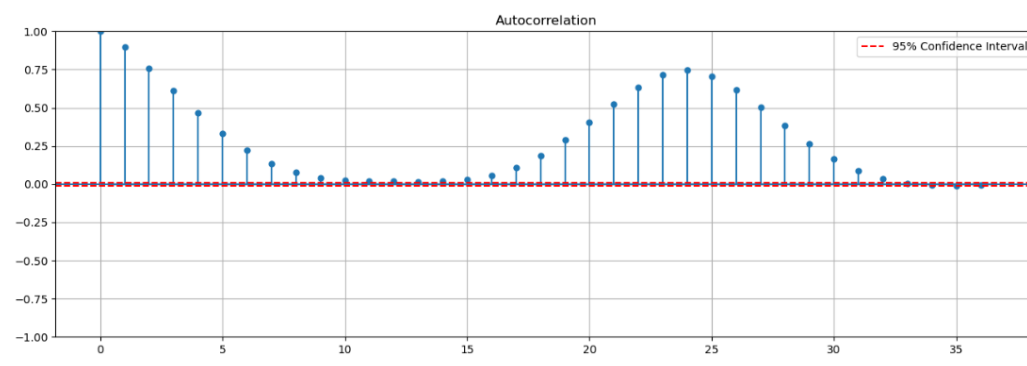
- D: درجه تفاضی گیری فصلی (SI)

- Q: درجه مدل میانگین متحرک فصلی (SMA)

- S: طول دوره فصلی

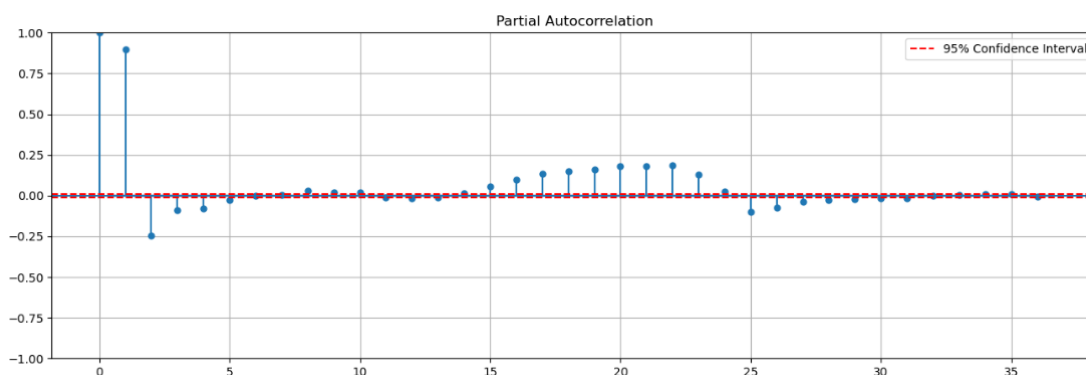
است.

با توجه به نمودار خود همبستگی، مقادیر خودهمبستگی در وقفه‌های اولیه (تا حدود 5-6 وقفه) نسبتاً بالا هستند و به تدریج کاهش می‌یابند. مقادیر خودهمبستگی به سرعت به صفر میل می‌کنند پس می‌توان نتیجه گرفت که داده‌های سری زمانی ایستا هستند.



شکل 14- نمودار خودهمبستگی

با توجه به نمودار خود همبستگی جزئی، مقادیر خودهمبستگی جزئی در تاخیر اول بالا است. در وقفه های دوم و سوم و چهارم کاهش می‌یابد تا در فاصله اطمینان قرار می‌گیرد. در نتیجه مدل ترکیبی اتورگرسیون میانگین متحرک فصلی چند متغیره پیشنهاد می‌شود.



شکل 15- نمودار خودهمبستگی جزئی

مدل اول:

در جدول زیر مدل  $SARIMAX(1, 0, 1) (1, 1, 1, 12)$  بر روی داده‌های آموزش برازش داده شده است و مدل پیشنهادی بر روی داده‌های آزمایش اجرا شده است و با کمک معیارهای ارزیابی مدل بررسی شده است. تفسیر نتایج به شرح زیر است:

**Log Likelihood** : مقدار  $-131817.380$  - نشان‌دهنده‌ی میزان برازش مدل به داده‌ها است. هرچه این مقدار به صفر نزدیک‌تر باشد، مدل بهتر است.

**AIC (Akaike Information Criterion)** : مقدار  $263648.76$  است. این معیار برای مقایسه مدل‌ها استفاده می‌شود و هرچه کمتر باشد، مدل بهتر است.

**BIC (Bayesian Information Criterion)** : مقدار  $263707.074$  است. این معیار نیز برای مقایسه مدل‌ها استفاده می‌شود و هرچه کمتر باشد، مدل بهتر است.

**Ljung-Box (Q)** : مقدار  $0.03$  و  $p\text{-value} = 0.85$  نشان‌دهنده‌ی این است که باقیمانده‌ها خودهمبستگی ندارند.

**Jarque-Bera (JB)** : مقدار  $247735.54$  و  $p\text{-value} = 0.00$  نشان‌دهنده‌ی این است که باقیمانده‌ها نرمال نیستند.

**Heteroskedasticity (H)** : مقدار  $1.01$  و  $p\text{-value} = 0.67$  نشان‌دهنده‌ی این است که ناهمسانی واریانس وجود ندارد.

میانگین مربعات خطا و مجذور میانگین مربعات خطا بزرگ هستند در حالی که  $R^2$  عدد بسیار کوچکی است. در زیر برخی از فرمول‌ها آورده شده است که در این روابط  $k$  تعداد پارامتر،  $N$  تعداد مشاهدات و  $L$  مقدار برآورد تابع ماکسیمم درست‌نمایی است.

$$\begin{aligned} RMSE &= \sqrt{\frac{\sum_{i=1}^N (X_i - \bar{X})^2}{N}} \\ AIC &= 2k - 2 \ln(\hat{l}) \\ BIC &= k \ln(n) - 2 \ln(\hat{l}) \end{aligned}$$

SARIMAX(1, 0, 1) (1, 1, 1, 12)	مدل
30667	تعداد مشاهدات آموزش
-131817.380	Log Likelihood
263648.760	AIC
263707.074	BIC
263667.451	HQIC
0.03	آماره لیونگ-باکس
0.85	سطح معنی داری آماره لیونگ-باکس
24729.54	آماره ژارک-برا
0.00	سطح معنی داری آماره ژارک-برا
1.01	Heteroskedasticity (ناهمسانی واریانس)
0.67	سطح معنی داری Heteroskedasticity
-0.10	چولگی
7.40	کشیدگی
13144	تعداد مشاهدات آزمایش
314.7216	MSE
17.7404	RMSE
0.0029	R2

جدول 6- شاخص های مدل SARIMAX(1, 0, 1) (1, 1, 1, 12)

جدول ضرایب مدل: (Coefficients)

دما: ضریب دما 0.0088 و p-value = 0.000 نشان دهنده‌ی این است که دما تأثیر معناداری بر غلظت ازن دارد.

سرعت وزش باد (wind\_space): ضریب -0.3025 و  $p\text{-value} = 0.000$  نشان‌دهنده‌ی این است که سرعت وزش باد تأثیر معناداری بر غلظت ازن دارد و این تأثیر منفی است.

AR (ar.L1): ضریب 0.6611 و  $p\text{-value} = 0.546$  نشان‌دهنده‌ی این است که تأثیر معناداری ندارد.

MA (ma.L1): ضریب -0.6634 و  $p\text{-value} = 0.543$  نشان‌دهنده‌ی این است که تأثیر معناداری ندارد.

AR (ar.S.L12): ضریب -0.0026 و  $p\text{-value} = 0.511$  نشان‌دهنده‌ی این است که تأثیر معناداری ندارد.

MA (ma.S.L12): ضریب 0.9995 و  $p\text{-value} = 0.000$  نشان‌دهنده‌ی این است که تأثیر معناداری دارد.

Sigma2: مقدار 317.2108 و  $p\text{-value} = 0.000$  نشان‌دهنده‌ی واریانس باقیمانده‌ها است که معنادار است.

	coef	std err	z	P> z	[0.025	0.975]
dama	0.0803	0.009	9.307	0.000	0.063	0.097
wind_space	-0.3025	0.051	-5.901	0.000	-0.403	-0.202
ar.L1	0.6611	1.095	0.604	0.546	-1.484	2.806
ma.L1	-0.6634	1.092	-0.608	0.543	-2.803	1.476
ar.S.L12	-0.0038	0.006	-0.658	0.511	-0.015	0.007
ma.S.L12	-0.9995	0.001	-1610.341	0.000	-1.001	-0.998
sigma2	317.2108	1.439	220.387	0.000	314.390	320.032

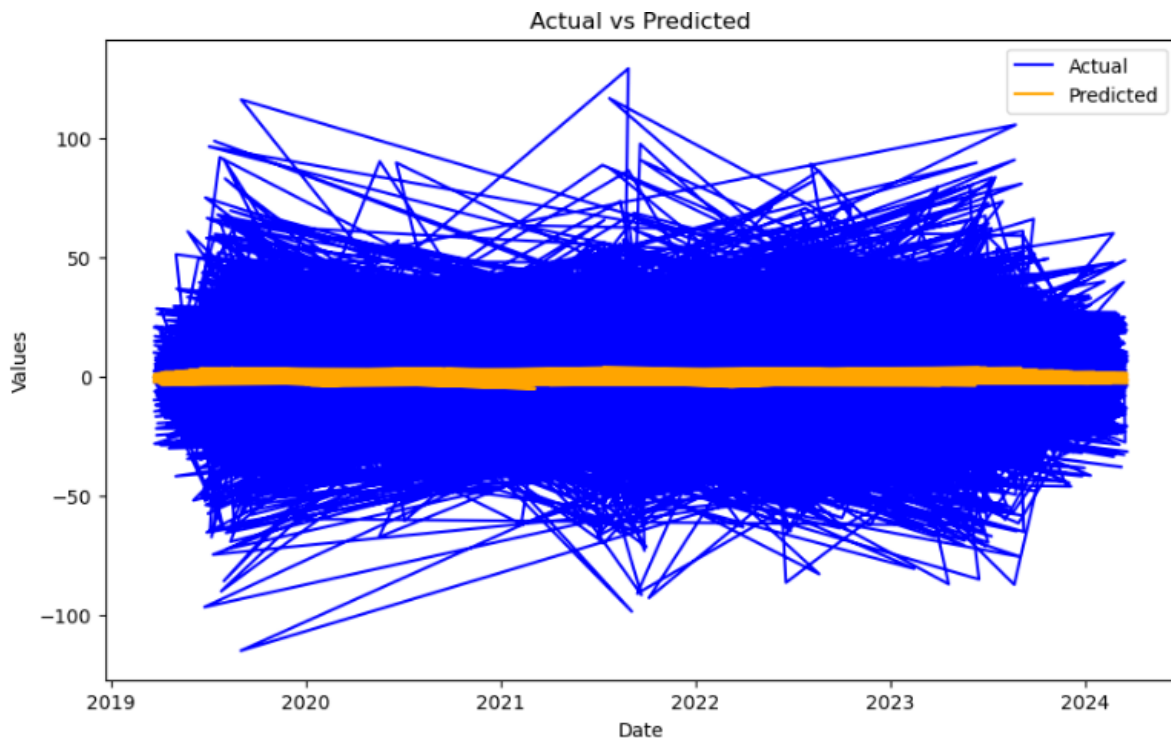
جدول 7- جدول ضرایب مدل SARIMAX(1, 0, 1) (1, 1, 1, 12)

برای بررسی بصری مدل و پیش‌بینی نمودار آن را در زیر داریم:

در این نمودار مقادیر واقعی مشاهدات در مقابل مقادیر پیش‌بینی شده کشیده شده است. خطوط آبی و نارنجی روند مشابهی ندارند و از هم فاصله زیادی دارند بنابراین مدل نتوانسته است روند کلی داده‌ها را به خوبی پیش‌بینی کند.

نتیجه‌گیری کلی:

با توجه به معیارهای ارزیابی مدل و نمودار مقادیر مشاهده شده در مقابل مقادیر پیش‌بینی شده، مدل SARIMAX(1, 0, 1) (1, 1, 1, 12) مدل مناسبی نمی‌تواند باشد و بهتر است مدل را با پارامترهای دیگری برازش دهیم یا از مدل‌های دیگری استفاده کنیم.



شکل 16- نمودار مقادیر پیش بینی و مقادیر واقعی مدل SARIMAX(1, 0, 1) (1, 1, 1, 12)

مدل دوم:

در جدول زیر مدل SARIMAX(2, 0, 2) (2, 1, 2, 12) بر روی داده‌های آموزش برازش داده شده است و مدل پیشنهادی بر روی داده‌های آزمایش اجرا شده است و با کمک معیارهای ارزیابی مدل بررسی شده است. تفسیر نتایج به شرح زیر است:

Log Likelihood : مقدار -131816.947- نشان‌دهنده‌ی میزان برازش مدل به داده‌ها است. هرچه این مقدار به صفر نزدیک‌تر باشد، مدل بهتر است.

AIC (Akaike Information Criterion) : مقدار 263655.894 است. این معیار برای مقایسه مدل‌ها استفاده می‌شود و هرچه کمتر باشد، مدل بهتر است.

BIC (Bayesian Information Criterion) : مقدار 263747.531 است. این معیار نیز برای مقایسه مدل‌ها استفاده می‌شود و هرچه کمتر باشد، مدل بهتر است.

Ljung-Box (Q) : مقدار 0.24 و  $p\text{-value} = 0.62$  نشان‌دهنده‌ی این است که باقیمانده‌ها خودهمبستگی ندارند.

Jarque-Bera (JB) : مقدار 24728.44 و  $p\text{-value} = 0.00$  نشان‌دهنده‌ی این است که باقیمانده‌ها نرمال نیستند.

Heteroskedasticity (H) : مقدار 1.01 و  $p\text{-value} = 0.65$  نشان دهنده‌ی این است که ناهمسانی واریانس وجود ندارد.

میانگین مربعات خطا و مجذور میانگین مربعات خطا بزرگ هستند در حالی که  $R^2$  عدد بسیار کوچکی است.

SARIMAX(2, 0, 2) (2, 1, 2, 12)	مدل
30667	تعداد مشاهدات آموزش
-131816.947	Log Likelihood
263655.894	AIC
263747.531	BIC
263685.267	HQIC
0.24	آماره لیونگ-باکس
0.62	سطح معنی داری آماره لیونگ-باکس
24728.44	آماره ژارک-برا
0.00	سطح معنی داری آماره ژارک-برا
1.01	Heteroskedasticity (ناهمسانی واریانس)
0.65	سطح معنی داری Heteroskedasticity
-0.10	چولگی
7.40	کشیدگی
13144	تعداد مشاهدات آزمایش
314.6761	MSE
0.0030	$R^2$

جدول 8- شاخص های مدل SARIMAX(2, 0, 2) (2, 1, 2, 12)

جدول ضرایب مدل: (Coefficients)

دما : ضریب 0.0802 و  $p\text{-value} = 0.000$  نشان‌دهنده‌ی این است که دما تأثیر معناداری بر متغیر وابسته دارد. این تأثیر مثبت است، یعنی با افزایش دما، غلظت ازن افزایش می‌یابد.

سرعت وزش باد : ضریب -0.3011 و  $p\text{-value} = 0.000$  نشان‌دهنده‌ی این است که سرعت وزش باد تأثیر معناداری بر متغیر وابسته دارد و این تأثیر منفی است، یعنی با افزایش سرعت وزش باد، غلظت ازن کاهش می‌یابد.

:AR (Autoregressive)

ar\_L1: ضریب -0.3022 و  $p\text{-value} = 0.963$  نشان‌دهنده‌ی این است که این جزء تأثیر معناداری ندارد.

ar\_L2: ضریب 0.2911 و  $p\text{-value} = 0.953$  نشان‌دهنده‌ی این است که این جزء تأثیر معناداری ندارد.

: MA (Moving Average)

ma\_L1: ضریب 0.2983 و  $p\text{-value} = 0.963$  نشان‌دهنده‌ی این است که این جزء تأثیر معناداری ندارد.

ma\_L2: ضریب -0.2885 و  $p\text{-value} = 0.953$  نشان‌دهنده‌ی این است که این جزء تأثیر معناداری ندارد.

: AR (Seasonal AR)

ar\_S\_L12: ضریب -0.9772 و  $p\text{-value} = 0.000$  نشان‌دهنده‌ی این است که این جزء تأثیر معناداری دارد.

ar\_S\_L24: ضریب -0.0035 و  $p\text{-value} = 0.581$  نشان‌دهنده‌ی این است که این جزء تأثیر معناداری ندارد.

: MA (Seasonal MA)

ma\_S\_L12: ضریب -0.0265 و  $p\text{-value} = 0.916$  نشان‌دهنده‌ی این است که این جزء تأثیر معناداری ندارد.

ma\_S\_L24: ضریب 0.9731 و  $p\text{-value} = 0.000$  نشان‌دهنده‌ی این است که این جزء تأثیر معناداری دارد.

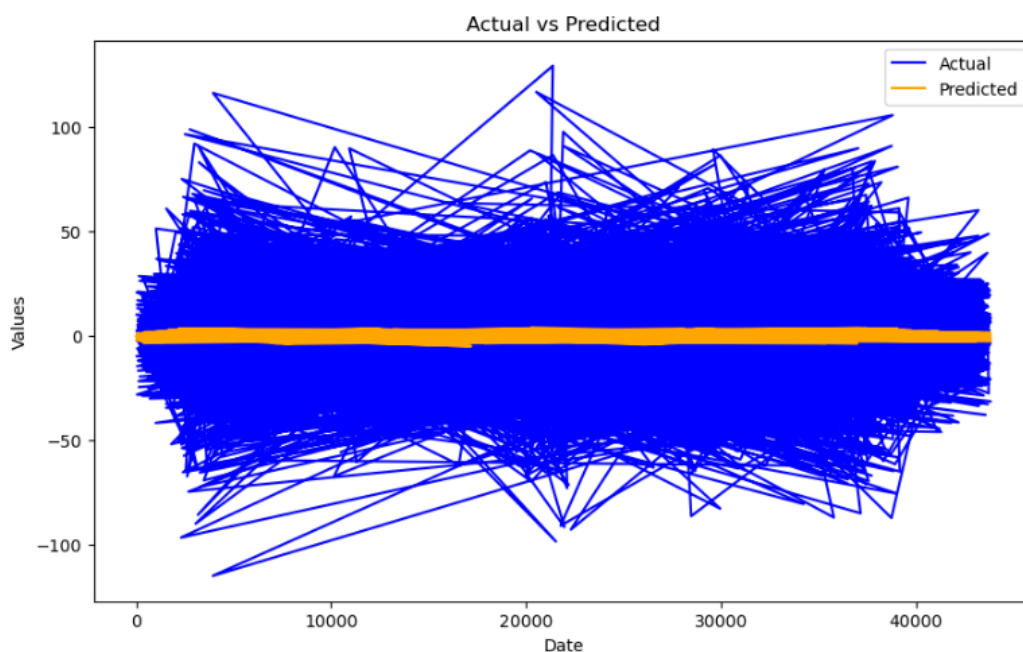
sigma2 : مقدار 317.1046 و  $p\text{-value} = 0.000$  نشان‌دهنده‌ی واریانس باقیمانده‌ها است که معنادار است.

	coef	std err	z	P> z	[0.025	0.975]
dama	0.0802	0.009	9.289	0.000	0.063	0.097
wind_space	-0.3011	0.051	-5.870	0.000	-0.402	-0.201
ar.L1	-0.3022	6.449	-0.047	0.963	-12.943	12.339
ar.L2	0.2911	4.900	0.059	0.953	-9.313	9.896
ma.L1	0.2983	6.451	0.046	0.963	-12.345	12.941
ma.L2	-0.2885	4.875	-0.059	0.953	-9.844	9.267
ar.S.L12	-0.9772	0.252	-3.874	0.000	-1.472	-0.483
ar.S.L24	-0.0035	0.006	-0.552	0.581	-0.016	0.009
ma.S.L12	-0.0265	0.252	-0.105	0.916	-0.520	0.467
ma.S.L24	-0.9731	0.252	-3.862	0.000	-1.467	-0.479
sigma2	317.1046	1.487	213.263	0.000	314.190	320.019

جدول 9- ضرایب مدل SARIMAX(2, 0, 2) (2, 1, 2, 12)

برای بررسی بصری مدل و پیش‌بینی نمودار آن را در زیر داریم:

در این نمودار مقادیر واقعی مشاهدات در مقابل مقادیر پیش‌بینی شده کشیده شده است. خطوط آبی و نارنجی روند مشابهی ندارند و از هم فاصله زیادی دارند بنابراین مدل نتوانسته است روند کلی داده‌ها را به خوبی پیش‌بینی کند.



شکل 17- نمودار مقادیر پیش‌بینی و مقادیر واقعی مدل SARIMAX(2, 0, 2) (2, 1, 2, 12)



نتیجه گیری کلی:

با توجه به معیارهای ارزیابی مدل و نمودار مقادیر مشاهده شده در مقابل مقادیر پیش بینی شده ، مدل  $SARIMAX(2, 0, 2) (2, 1, 2, 12)$  مدل مناسبی نمیتواند باشد و بهتر است مدل را با پارامترهای دیگری برازش دهیم یا از مدل های دیگری استفاده کنیم.

مدل سوم:

مدل  $SARIMAX(3, 0, 3) \times (1, 1, [1], 12)$  را مانند مدل اول و دوم به داده ها برازش دادیم و نتایج مشابهی داشت. پس به نظر می رسد مدل  $SARIMAX$  در کل مناسب نیست.

مدل چهارم:

### VAR (Vector Autoregressive)

مدل VAR یک مدل آماری است که برای تحلیل رابطه های پویا بین چندین سری زمانی استفاده می شود. این مدل می تواند برای پیش بینی و بررسی تأثیرات متقابل بین متغیرها مفید باشد. در این مدل داده ها را به دو بخش آموزش و آزمایش تقسیم کرده ایم سپس مدل را بر روی داده های آموزش برازش دادیم و با استفاده از معیارهای ارزیابی مدل آن را بررسی کردیم. این معیارها به تعیین تعداد تأخیرهای بهینه برای مدل کمک می کنند. داده ها با فرکانس ساعتی (هر ساعت یک داده) آموزش داده می شوند. مدل با ۲۷ تأخیر (lag) آموزش داده می شود. این تعداد تأخیر بر اساس حداقل مقدار AIC و BIC انتخاب شده است.

VAR Order Selection (* highlights the minimums)				
	AIC	BIC	FPE	HQIC
0	12.88	12.88	3.911e+05	12.88
1	7.538	7.541	1878.	7.539
2	7.214	7.220	1359.	7.216
3	7.116	7.124	1231.	7.118
4	7.087	7.098	1196.	7.091
5	7.076	7.089	1183.	7.080
6	7.068	7.084	1174.	7.073
7	7.060	7.078	1165.	7.066
8	7.046	7.067	1149.	7.053
9	7.028	7.050	1127.	7.035
10	7.002	7.027	1099.	7.010
11	6.973	7.001	1067.	6.982
12	6.939	6.969	1032.	6.949
13	6.905	6.938	997.5	6.916
14	6.876	6.911	968.7	6.887
15	6.853	6.890	946.5	6.865
16	6.827	6.867	922.2	6.840
17	6.791	6.833	889.8	6.805
18	6.748	6.793	852.2	6.762
19	6.700	6.748	812.8	6.716
20	6.632	6.681	758.6	6.647
21	6.558	6.610	704.7	6.575
22	6.488	6.543	657.5	6.506
23	6.450	6.507	632.5	6.468
24	6.446	6.506	630.5	6.466
25	6.384	6.446	592.6	6.404
26	6.348	6.413	571.5	6.369
27	6.341	6.408*	567.4	6.363
28	6.340	6.410	566.9	6.362
29	6.339	6.411	566.3	6.362
30	6.338*	6.412	565.8*	6.362*

جدول 10- شاخص های برازش مدل VAR

با توجه به جدول زیر مدل با استفاده از روش حداقل مربعات به برآورد پارامترها پرداخته است. نسبت به مدل SARIMAX ، مدل VAR معیارهای AIC , BIC کمتری را نشان میدهد.

Summary of Regression Results			
=====			
Model:	VAR		
Method:	OLS		
Date:	Thu, 02, Jan, 2025		
Time:	12:08:50		
-----			
No. of Equations:	3.00000	BIC:	6.41494
Nobs:	30649.0	HQIC:	6.36951
Log likelihood:	-227502.	FPE:	571.393
AIC:	6.34808	Det(Omega_mle):	566.831
-----			

جدول 11- خلاصه مدل VAR

به منظور تعیین ضرایب معنادار برای هر کدام از متغیرها از آزمون تی-تست استفاده کردیم و با توجه به سطح معنی داری میتوان فهمید که کدام یک از ضرایب معنادار هستند و باید در مدل باقی بمانند. کمترین مقدار BIC در تاخیر 27 ام است پس تا تاخیر 27 آزمون تی و براورد ضرایب بررسی شده است. ضرایبی که سطح معناداری آنها کمتر از 0.05 هستند در مدل باقی میمانند و بقیه از مدل حذف می‌شوند.

#### Results for equation O3 (ppb)

=====				
===				
	coefficient	std. error	t-stat	prob
-----				
const	0.355165	0.190848	1.861	0.063
L1.O3 (ppb)	0.859879	0.005714	150.498	0.000
L1.dama	0.254060	0.040510	6.272	0.000
L1.wind_space	-0.093556	0.034184	-2.737	0.006
L2.O3 (ppb)	-0.091582	0.007538	-12.149	0.000
L2.dama	0.300665	0.051871	5.796	0.000
L2.wind_space	0.161952	0.037832	4.281	0.000
L3.O3 (ppb)	0.011961	0.007551	1.584	0.113
L3.dama	-0.141371	0.051888	-2.725	0.006
L3.wind_space	0.205789	0.038061	5.407	0.000
L4.O3 (ppb)	0.005729	0.007524	0.761	0.446
L4.dama	-0.158436	0.051388	-3.083	0.002
L4.wind_space	-0.048481	0.038094	-1.273	0.203
L5.O3 (ppb)	-0.022090	0.007509	-2.942	0.003
L5.dama	-0.004445	0.051036	-0.087	0.931
L5.wind_space	-0.045320	0.038093	-1.190	0.234
L6.O3 (ppb)	0.013922	0.007504	1.855	0.064
L6.dama	-0.054081	0.050984	-1.061	0.289

L6.wind_space	-0.051959	0.038068	-1.365	0.172
L7.O3 (ppb)	-0.025462	0.007505	-3.393	0.001
L7.dama	0.005407	0.050965	0.106	0.916
L7.wind_space	-0.013315	0.038057	-0.350	0.726
L8.O3 (ppb)	0.008141	0.007506	1.085	0.278
L8.dama	-0.082053	0.050960	-1.610	0.107
L8.wind_space	0.107270	0.038060	2.818	0.005
L9.O3 (ppb)	-0.010437	0.007507	-1.390	0.164
L9.dama	0.004096	0.050957	0.080	0.936
L9.wind_space	0.015048	0.038048	0.396	0.692
L10.O3 (ppb)	0.019050	0.007507	2.538	0.011
L10.dama	0.014394	0.050958	0.282	0.778
L10.wind_space	0.045313	0.038049	1.191	0.234
L11.O3 (ppb)	0.014974	0.007507	1.995	0.046
L11.dama	-0.071617	0.050954	-1.406	0.160
L11.wind_space	-0.042718	0.038052	-1.123	0.262
L12.O3 (ppb)	-0.002835	0.007507	-0.378	0.706
L12.dama	-0.011747	0.050959	-0.231	0.818
L12.wind_space	0.116969	0.038050	3.074	0.002
L13.O3 (ppb)	0.008142	0.007507	1.085	0.278
L13.dama	0.020167	0.050959	0.396	0.692
L13.wind_space	-0.083038	0.038057	-2.182	0.029
L14.O3 (ppb)	-0.005406	0.007508	-0.720	0.472
L14.dama	-0.065907	0.050958	-1.293	0.196
L14.wind_space	-0.052860	0.038059	-1.389	0.165
L15.O3 (ppb)	-0.019359	0.007507	-2.579	0.010
L15.dama	0.002285	0.050953	0.045	0.964

L15.wind_space	-0.029974	0.038059	-0.788	0.431
L16.O3 (ppb)	-0.006025	0.007507	-0.803	0.422
L16.dama	0.146972	0.050951	2.885	0.004
L16.wind_space	0.040307	0.038055	1.059	0.290
L17.O3 (ppb)	-0.006671	0.007507	-0.889	0.374
L17.dama	0.122497	0.050952	2.404	0.016
L17.wind_space	0.058619	0.038057	1.540	0.123
L18.O3 (ppb)	0.015930	0.007506	2.122	0.034
L18.dama	-0.125613	0.050951	-2.465	0.014
L18.wind_space	0.016178	0.038058	0.425	0.671
L19.O3 (ppb)	-0.004297	0.007506	-0.572	0.567
L19.dama	0.060933	0.050948	1.196	0.232
L19.wind_space	0.060790	0.038054	1.597	0.110
L20.O3 (ppb)	0.025939	0.007507	3.455	0.001
L20.dama	-0.024684	0.050944	-0.485	0.628
L20.wind_space	-0.024001	0.038054	-0.631	0.528
L21.O3 (ppb)	0.018284	0.007507	2.436	0.015
L21.dama	-0.061656	0.050945	-1.210	0.226
L21.wind_space	-0.016834	0.038051	-0.442	0.658
L22.O3 (ppb)	0.051987	0.007508	6.924	0.000
L22.dama	0.016260	0.050952	0.319	0.750
L22.wind_space	0.081396	0.038064	2.138	0.032
L23.O3 (ppb)	0.080316	0.007513	10.690	0.000
L23.dama	-0.227887	0.050999	-4.468	0.000
L23.wind_space	0.025385	0.038101	0.666	0.505
L24.O3 (ppb)	0.112991	0.007528	15.009	0.000
L24.dama	-0.050457	0.051359	-0.982	0.326

L24.wind_space	0.025851	0.038103	0.678	0.497
L25.O3 (ppb)	-0.052943	0.007553	-7.009	0.000
L25.dama	-0.057696	0.051853	-1.113	0.266
L25.wind_space	-0.027420	0.038083	-0.720	0.472
L26.O3 (ppb)	-0.021584	0.007538	-2.864	0.004
L26.dama	0.135612	0.051869	2.615	0.009
L26.wind_space	-0.084100	0.037864	-2.221	0.026
L27.O3 (ppb)	-0.038283	0.005698	-6.719	0.000
L27.dama	0.108739	0.040762	2.668	0.008
L27.wind_space	-0.110842	0.034151	-3.246	0.001

===

جدول 12- برآورد ضرایب آلاینده‌ی ازن

Results for equation dama

===

	coefficient	std. error	t-stat	prob
const	0.363294	0.026903	13.504	0.000
L1.O3 (ppb)	-0.002374	0.000805	-2.948	0.003
L1.dama	0.806701	0.005711	141.263	0.000
L1.wind_space	-0.009957	0.004819	-2.066	0.039
L2.O3 (ppb)	-0.001494	0.001063	-1.406	0.160
L2.dama	0.065260	0.007312	8.925	0.000
L2.wind_space	-0.006418	0.005333	-1.203	0.229
L3.O3 (ppb)	0.002097	0.001064	1.970	0.049
L3.dama	-0.001675	0.007315	-0.229	0.819
L3.wind_space	-0.009910	0.005365	-1.847	0.065

L4.O3 (ppb)	0.000048	0.001061	0.046	0.964
L4.dama	-0.007613	0.007244	-1.051	0.293
L4.wind_space	-0.006411	0.005370	-1.194	0.233
L5.O3 (ppb)	-0.000380	0.001059	-0.359	0.719
L5.dama	0.006387	0.007194	0.888	0.375
L5.wind_space	-0.007897	0.005370	-1.471	0.141
L6.O3 (ppb)	0.000057	0.001058	0.054	0.957
L6.dama	0.001596	0.007187	0.222	0.824
L6.wind_space	0.009817	0.005366	1.829	0.067
L7.O3 (ppb)	-0.001539	0.001058	-1.455	0.146
L7.dama	0.003213	0.007184	0.447	0.655
L7.wind_space	-0.004495	0.005365	-0.838	0.402
L8.O3 (ppb)	0.002018	0.001058	1.907	0.057
L8.dama	-0.011109	0.007184	-1.546	0.122
L8.wind_space	-0.005005	0.005365	-0.933	0.351
L9.O3 (ppb)	-0.003040	0.001058	-2.873	0.004
L9.dama	-0.000192	0.007183	-0.027	0.979
L9.wind_space	-0.000564	0.005364	-0.105	0.916
L10.O3 (ppb)	-0.001810	0.001058	-1.711	0.087
L10.dama	0.001633	0.007183	0.227	0.820
L10.wind_space	-0.018021	0.005364	-3.360	0.001
L11.O3 (ppb)	0.001643	0.001058	1.553	0.121
L11.dama	-0.017411	0.007183	-2.424	0.015
L11.wind_space	-0.007507	0.005364	-1.399	0.162
L12.O3 (ppb)	0.000917	0.001058	0.866	0.386
L12.dama	-0.008126	0.007184	-1.131	0.258
L12.wind_space	-0.008319	0.005364	-1.551	0.121

L13.O3 (ppb)	-0.002960	0.001058	-2.797	0.005
L13.dama	0.013745	0.007184	1.913	0.056
L13.wind_space	-0.003732	0.005365	-0.696	0.487
L14.O3 (ppb)	-0.000010	0.001058	-0.009	0.992
L14.dama	0.013318	0.007184	1.854	0.064
L14.wind_space	0.002252	0.005365	0.420	0.675
L15.O3 (ppb)	0.001079	0.001058	1.020	0.308
L15.dama	-0.002882	0.007183	-0.401	0.688
L15.wind_space	-0.003865	0.005365	-0.720	0.471
L16.O3 (ppb)	-0.001647	0.001058	-1.556	0.120
L16.dama	0.000552	0.007182	0.077	0.939
L16.wind_space	-0.005433	0.005365	-1.013	0.311
L17.O3 (ppb)	0.001102	0.001058	1.041	0.298
L17.dama	0.012101	0.007183	1.685	0.092
L17.wind_space	-0.008913	0.005365	-1.661	0.097
L18.O3 (ppb)	0.000931	0.001058	0.880	0.379
L18.dama	0.005561	0.007183	0.774	0.439
L18.wind_space	-0.007560	0.005365	-1.409	0.159
L19.O3 (ppb)	0.000540	0.001058	0.510	0.610
L19.dama	-0.018899	0.007182	-2.631	0.009
L19.wind_space	-0.001411	0.005364	-0.263	0.793
L20.O3 (ppb)	0.001432	0.001058	1.353	0.176
L20.dama	0.005016	0.007181	0.698	0.485
L20.wind_space	0.000671	0.005364	0.125	0.900
L21.O3 (ppb)	0.002697	0.001058	2.548	0.011
L21.dama	0.020583	0.007182	2.866	0.004
L21.wind_space	-0.001505	0.005364	-0.281	0.779



L22.O3 (ppb)	-0.000293	0.001058	-0.277	0.782
L22.dama	0.055265	0.007183	7.694	0.000
L22.wind_space	-0.015961	0.005366	-2.975	0.003
L23.O3 (ppb)	-0.004403	0.001059	-4.157	0.000
L23.dama	0.145865	0.007189	20.289	0.000
L23.wind_space	-0.007605	0.005371	-1.416	0.157
L24.O3 (ppb)	-0.002421	0.001061	-2.281	0.023
L24.dama	0.175465	0.007240	24.235	0.000
L24.wind_space	0.003099	0.005371	0.577	0.564
L25.O3 (ppb)	0.001359	0.001065	1.276	0.202
L25.dama	-0.071847	0.007310	-9.829	0.000
L25.wind_space	0.011709	0.005368	2.181	0.029
L26.O3 (ppb)	0.001461	0.001063	1.375	0.169
L26.dama	-0.119726	0.007312	-16.374	0.000
L26.wind_space	0.002015	0.005338	0.377	0.706
L27.O3 (ppb)	0.002439	0.000803	3.037	0.002
L27.dama	-0.069963	0.005746	-12.176	0.000
L27.wind_space	-0.000511	0.004814	-0.106	0.915

جدول 13- برآورد ضرایب دما

Results for equation wind\_space

=====				
===				
	coefficient	std. error	t-stat	prob
-----				
const	0.399355	0.031953	12.498	0.000
L1.O3 (ppb)	0.000531	0.000957	0.555	0.579

L1.dama	0.042004	0.006782	6.193	0.000
L1.wind_space	0.473418	0.005723	82.718	0.000
L2.O3 (ppb)	0.000805	0.001262	0.638	0.524
L2.dama	0.003335	0.008685	0.384	0.701
L2.wind_space	0.117958	0.006334	18.623	0.000
L3.O3 (ppb)	0.000225	0.001264	0.178	0.859
L3.dama	0.012109	0.008687	1.394	0.163
L3.wind_space	0.042255	0.006372	6.631	0.000
L4.O3 (ppb)	-0.001577	0.001260	-1.252	0.211
L4.dama	0.014724	0.008604	1.711	0.087
L4.wind_space	0.026422	0.006378	4.143	0.000
L5.O3 (ppb)	0.000919	0.001257	0.731	0.465
L5.dama	-0.000316	0.008545	-0.037	0.970
L5.wind_space	0.010289	0.006378	1.613	0.107
L6.O3 (ppb)	-0.000169	0.001256	-0.134	0.893
L6.dama	-0.008238	0.008536	-0.965	0.335
L6.wind_space	0.016569	0.006374	2.600	0.009
L7.O3 (ppb)	0.002764	0.001256	2.200	0.028
L7.dama	-0.012639	0.008533	-1.481	0.139
L7.wind_space	-0.010665	0.006372	-1.674	0.094
L8.O3 (ppb)	-0.002749	0.001257	-2.187	0.029
L8.dama	-0.005906	0.008532	-0.692	0.489
L8.wind_space	0.003750	0.006372	0.589	0.556
L9.O3 (ppb)	-0.001600	0.001257	-1.273	0.203
L9.dama	0.012910	0.008532	1.513	0.130
L9.wind_space	-0.010783	0.006370	-1.693	0.091
L10.O3 (ppb)	-0.001075	0.001257	-0.855	0.393

L10.dama	0.003891	0.008532	0.456	0.648
L10.wind_space	-0.004959	0.006370	-0.778	0.436
L11.O3 (ppb)	0.000276	0.001257	0.220	0.826
L11.dama	-0.009301	0.008531	-1.090	0.276
L11.wind_space	0.003462	0.006371	0.543	0.587
L12.O3 (ppb)	0.000967	0.001257	0.769	0.442
L12.dama	0.000809	0.008532	0.095	0.924
L12.wind_space	0.018202	0.006370	2.857	0.004
L13.O3 (ppb)	0.001090	0.001257	0.868	0.386
L13.dama	-0.003349	0.008532	-0.393	0.695
L13.wind_space	-0.008661	0.006372	-1.359	0.174
L14.O3 (ppb)	-0.001652	0.001257	-1.314	0.189
L14.dama	-0.007447	0.008532	-0.873	0.383
L14.wind_space	0.002998	0.006372	0.470	0.638
L15.O3 (ppb)	0.000459	0.001257	0.365	0.715
L15.dama	0.001155	0.008531	0.135	0.892
L15.wind_space	0.007531	0.006372	1.182	0.237
L16.O3 (ppb)	-0.001244	0.001257	-0.990	0.322
L16.dama	0.013337	0.008530	1.563	0.118
L16.wind_space	0.008012	0.006371	1.258	0.209
L17.O3 (ppb)	0.001291	0.001257	1.027	0.304
L17.dama	0.000037	0.008531	0.004	0.997
L17.wind_space	0.002610	0.006372	0.410	0.682
L18.O3 (ppb)	0.000719	0.001257	0.572	0.567
L18.dama	-0.016600	0.008531	-1.946	0.052
L18.wind_space	0.004200	0.006372	0.659	0.510
L19.O3 (ppb)	-0.003828	0.001257	-3.046	0.002

L19.dama	0.018914	0.008530	2.217	0.027
L19.wind_space	0.018157	0.006371	2.850	0.004
L20.O3 (ppb)	0.001651	0.001257	1.314	0.189
L20.dama	-0.004467	0.008529	-0.524	0.600
L20.wind_space	0.005733	0.006371	0.900	0.368
L21.O3 (ppb)	-0.001044	0.001257	-0.831	0.406
L21.dama	-0.009788	0.008529	-1.148	0.251
L21.wind_space	0.032528	0.006371	5.106	0.000
L22.O3 (ppb)	-0.000472	0.001257	-0.376	0.707
L22.dama	0.000685	0.008531	0.080	0.936
L22.wind_space	0.042779	0.006373	6.713	0.000
L23.O3 (ppb)	-0.000924	0.001258	-0.735	0.462
L23.dama	0.004061	0.008539	0.476	0.634
L23.wind_space	0.031024	0.006379	4.863	0.000
L24.O3 (ppb)	0.001171	0.001260	0.929	0.353
L24.dama	-0.010686	0.008599	-1.243	0.214
L24.wind_space	0.027202	0.006379	4.264	0.000
L25.O3 (ppb)	0.000464	0.001265	0.367	0.714
L25.dama	0.002996	0.008682	0.345	0.730
L25.wind_space	0.011360	0.006376	1.782	0.075
L26.O3 (ppb)	0.001988	0.001262	1.576	0.115
L26.dama	-0.016532	0.008684	-1.904	0.057
L26.wind_space	-0.008604	0.006339	-1.357	0.175
L27.O3 (ppb)	-0.003135	0.000954	-3.286	0.001
L27.dama	-0.015700	0.006825	-2.301	0.021
L27.wind_space	-0.011822	0.005718	-2.068	0.039

جدول 13- برآورد ضرایب سرعت وزش باد

ماتریس واریانس-کواریانس باقیمانده‌ها :

باقیمانده‌ها تفاوت بین مقادیر واقعی و مقادیر پیش‌بینی شده توسط مدل هستند. اگر باقیمانده‌ها با هم همبستگی نداشته باشند، نشان‌دهنده‌ی این است که مدل به خوبی برازش شده است. نتایج حاکی از آن است که بین باقیمانده‌ها به صورت دو به دو همبستگی قوی وجود ندارد بنابراین مدل VAR مدل مناسبی برای برازش به داده‌ها است.

	آلایندهی ازن	دما	سرعت وزش باد
آلایندهی ازن	1.000000	0.005767	-0.002308
دما	0.005767	1.000000	0.043634
سرعت وزش باد	-0.002308	0.043634	1.000000

جدول 14- ماتریس واریانس کواریانس متغیرها

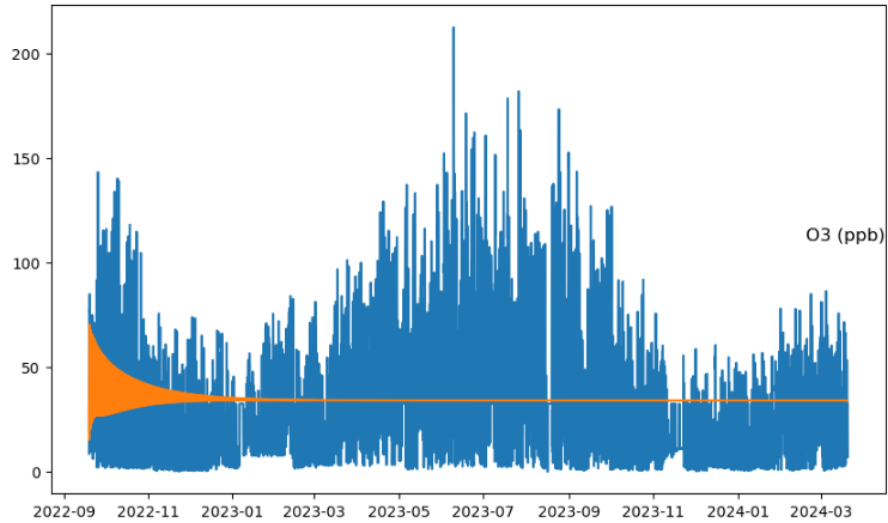
اکنون که مدل روی داده‌های آموزش به خوبی برازش داده شده است بر روی داده‌های آزمایش آن را اجرا میکنیم.

	O3 (ppb)	dama	wind_space
Gregorian_DateTime			
2022-09-19 04:00:00	29.973856	15.513005	1.829291
2022-09-19 05:00:00	32.419147	14.523638	1.922698
2022-09-19 06:00:00	31.339366	14.686278	1.990082
2022-09-19 07:00:00	24.066403	16.531154	1.995000
2022-09-19 08:00:00	16.968452	20.347389	1.909526
...	...	...	...
2024-03-19 19:00:00	33.857646	17.870606	2.935958
2024-03-19 20:00:00	33.857646	17.870606	2.935958
2024-03-19 21:00:00	33.857646	17.870606	2.935958
2024-03-19 22:00:00	33.857646	17.870606	2.935958
2024-03-19 23:00:00	33.857646	17.870606	2.935958

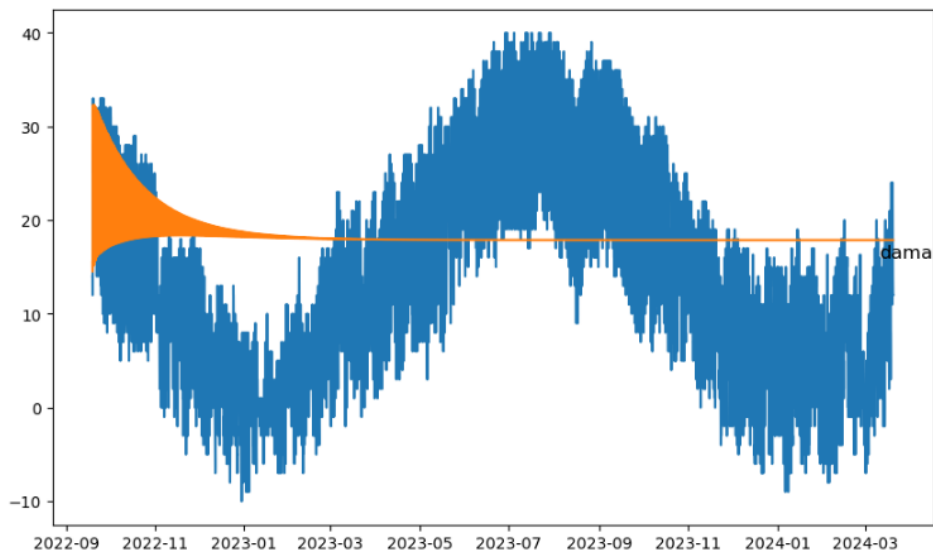
13148 rows × 3 columns

جدول 15- برازش مدل روی داده‌های آزمایش

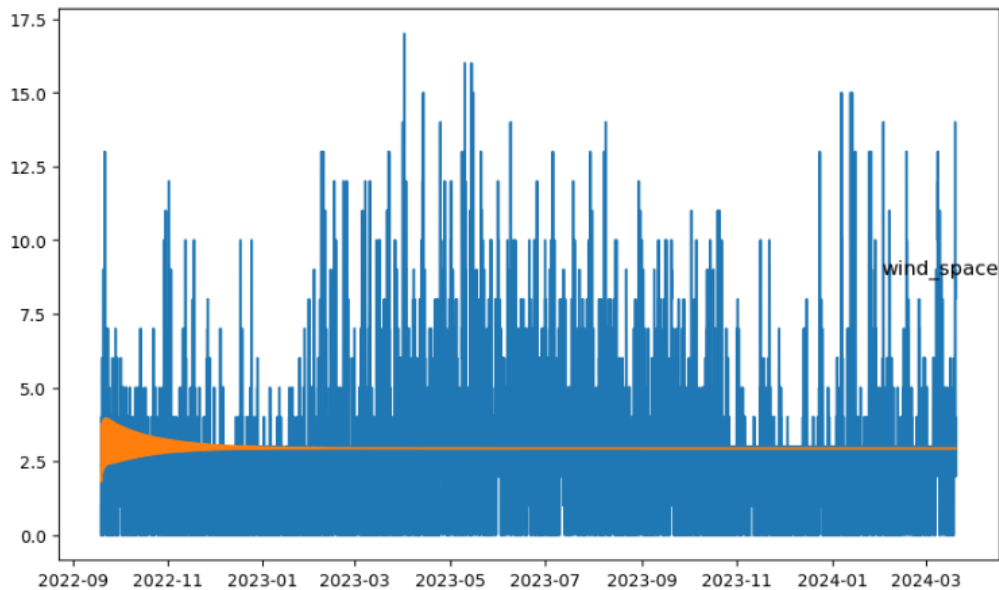
نمودار مقادیر پیش بینی شده را در مقابل مقادیر برازش داده شده در شکل های زیر داریم. با توجه به نمودارها مدل VAR از مدل SARIMAX عملکرد بهتری داشته است ولی همچنان تفاوت زیادی بین مقادیر مشاهده شده و مقادیر برازش داده شده مشاهده میشود.



شکل 18-نمودار مقادیر واقعی در مقابل مقادیر پیش بینی شده آلاینده ی ازن



شکل 19-نمودار مقادیر واقعی در مقابل مقادیر پیش بینی شده دما



شکل 19- نمودار مقادیر واقعی در مقابل مقادیر پیش بینی شده سرعت وزش باد

مدل پنجم:

#### مدل LSTM (Long Short-Term Memory)

مدل LSTM یک نوع شبکه عصبی بازگشتی (Recurrent Neural Network) است که برای پردازش داده‌های ترتیبی مانند سری‌های زمانی، داده‌های متنی و داده‌های صوتی به کار می‌رود. یکی از مشهورترین مدل‌ها برای بررسی سری‌های زمانی چند متغیره استفاده از مدل شبکه‌های عصبی در برازش مدل است. برای اجرای این مدل مراحل زیر دنبال می‌شود:

- نرمال‌سازی داده‌ها: داده‌ها با روش‌های نرمال‌سازی مانند MinMaxScaler یا StandardScaler مقیاس‌بندی می‌شوند زیرا مدل LSTM با داده‌های نرمال‌سازی شده بهتر عمل می‌کند.
- ایجاد توالی‌ها: داده‌ها به توالی‌هایی مناسب مدل LSTM تبدیل می‌شود. برای مثال، در پیش‌بینی سری زمانی، پنجره‌ای از داده‌های گذشته به عنوان ورودی و نقطه بعدی به عنوان خروجی تعریف می‌شود.
- تقسیم داده‌ها به مجموعه‌های آموزشی و آزمایشی و اعتبارسنجی: داده‌ها را به سه بخش آموزشی (Training) و آزمایشی (Testing) و اعتبارسنجی (validation) تقسیم می‌شود.
- طراحی مدل: آرگومان‌های مدل انتخاب می‌شود و مدل طراحی می‌گردد.

برآزش مدل: مدل بر روی داده‌های آموزشی، آموزش میبندد و از داده‌های اعتبارسنجی برای ارزیابی عملکرد استفاده می‌گردد

ارزیابی مدل: عملکرد مدل با داده‌های آزمایشی سنجیده می‌شود. از معیارهایی مانند MSE ، MAE یا دقت برای ارزیابی استفاده می‌گردد

پیش‌بینی با مدل: از مدل برای پیش‌بینی مقادیر جدید استفاده کنید. داده‌های نرمال شده را معکوس کرده و به شکل اولیه باز خواهند گشت.

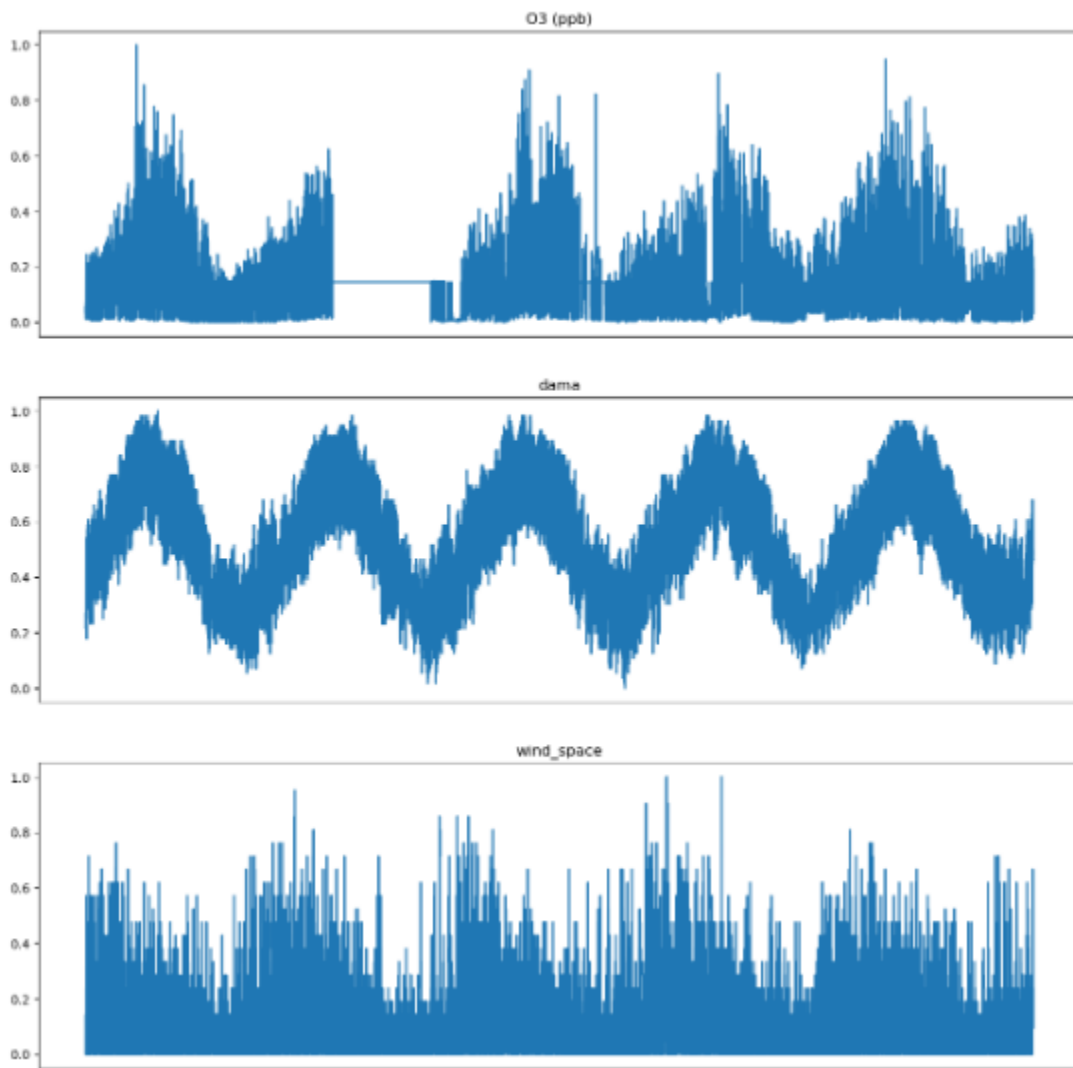
در زیر داده‌ها پس از نرمال سازی مشاهده می‌گردد.

	O3 (ppb)	dama	wind_space
Gregorian_DateTime			
2019-03-21 00:00:00	0.034802	0.267857	0.000000
2019-03-21 01:00:00	0.058792	0.267857	0.000000
2019-03-21 02:00:00	0.017334	0.250000	0.142857
2019-03-21 03:00:00	0.011347	0.214286	0.142857
2019-03-21 04:00:00	0.033238	0.232143	0.095238

جدول 16- داده‌های نرمال شده



نمودار داده‌های نرمال شده در زیر آورده شده است. مقیاس داده‌ها از صفر تا یک تغییر می‌کنند.



شکل 20- سری زمانی داده‌های نرمال شده

در زیر بخشی از نتایج آموزش مدل مشاهده می‌شود.

در این مدل از الگوریتم ADAM و تابع هزینه MSE برای مدلسازی استفاده شده است. تعداد Epoch = 100 دوره برای آموزش است. هر دوره یک بار عبور کامل از داده‌های آموزشی را نشان می‌دهد.

مقادیر `val_RootMeanSquaredError` و `val_loss` عملکرد مدل روی مجموعه داده اعتبارسنجی را نشان می‌دهند و به بررسی تعمیم‌پذیری مدل کمک می‌کنند.

معیارهای اعتبارسنجی به آرامی در حال بهبود هستند که نشان‌دهنده روند مثبت آموزش است.

نشانه‌ای از بیش‌برازش (Overfitting) دیده نمی‌شود، چرا که Loss اعتبارسنجی افزایش پیدا نکرده است.

```
Epoch 1/100
9336/9336 — 423s 44ms/step - RootMeanSquaredError: 0.0979 - loss: 0.0099 - val_RootMeanSquaredError: 0.0586 - val_loss: 0.0034
Epoch 2/100
9336/9336 — 402s 43ms/step - RootMeanSquaredError: 0.0678 - loss: 0.0046 - val_RootMeanSquaredError: 0.0573 - val_loss: 0.0033
Epoch 3/100
9336/9336 — 414s 44ms/step - RootMeanSquaredError: 0.0654 - loss: 0.0043 - val_RootMeanSquaredError: 0.0591 - val_loss: 0.0035
Epoch 4/100
9336/9336 — 448s 45ms/step - RootMeanSquaredError: 0.0642 - loss: 0.0041 - val_RootMeanSquaredError: 0.0577 - val_loss: 0.0033
Epoch 5/100
9336/9336 — 409s 44ms/step - RootMeanSquaredError: 0.0630 - loss: 0.0040 - val_RootMeanSquaredError: 0.0575 - val_loss: 0.0033
Epoch 6/100
9336/9336 — 414s 44ms/step - RootMeanSquaredError: 0.0628 - loss: 0.0039 - val_RootMeanSquaredError: 0.0568 - val_loss: 0.0032
Epoch 7/100
9336/9336 — 454s 46ms/step - RootMeanSquaredError: 0.0627 - loss: 0.0039 - val_RootMeanSquaredError: 0.0560 - val_loss: 0.0031
Epoch 8/100
9336/9336 — 426s 46ms/step - RootMeanSquaredError: 0.0615 - loss: 0.0038 - val_RootMeanSquaredError: 0.0556 - val_loss: 0.0031
Epoch 9/100
9336/9336 — 408s 44ms/step - RootMeanSquaredError: 0.0615 - loss: 0.0038 - val_RootMeanSquaredError: 0.0547 - val_loss: 0.0030
Epoch 10/100
9336/9336 — 397s 43ms/step - RootMeanSquaredError: 0.0614 - loss: 0.0038 - val_RootMeanSquaredError: 0.0554 - val_loss: 0.0031
Epoch 11/100
9336/9336 — 399s 43ms/step - RootMeanSquaredError: 0.0614 - loss: 0.0038 - val_RootMeanSquaredError: 0.0561 - val_loss: 0.0031
Epoch 12/100
9336/9336 — 18909s 2s/step - RootMeanSquaredError: 0.0602 - loss: 0.0036 - val_RootMeanSquaredError: 0.0553 - val_loss: 0.0031
```

جدول 17- برازش مدل بر روی داده‌های آموزش

بر روی داده‌های آزمایش نرمال شده مدل اجرا شده است. بخشی از نتایج به صورت زیر است:

```
array([[0.1392649 , 0.8611276 , 0.11658723],
       [0.04952632, 0.7968938 , 0.09444292],
       [0.1751412 , 0.49054793, 0.11400533],
       [0.01018026, 0.18735516, 0.08746219],
       [0.22980258, 0.64865756, 0.16613401],
       [0.08655792, 0.66786736, 0.15784419],
       [0.13728553, 0.6482172 , 0.12471095],
       [0.3487121 , 0.94344985, 0.20388082],
       [0.04105264, 0.579897 , 0.13227576],
       [0.13683277, 0.47267652, 0.10514687]], dtype=float32)
```

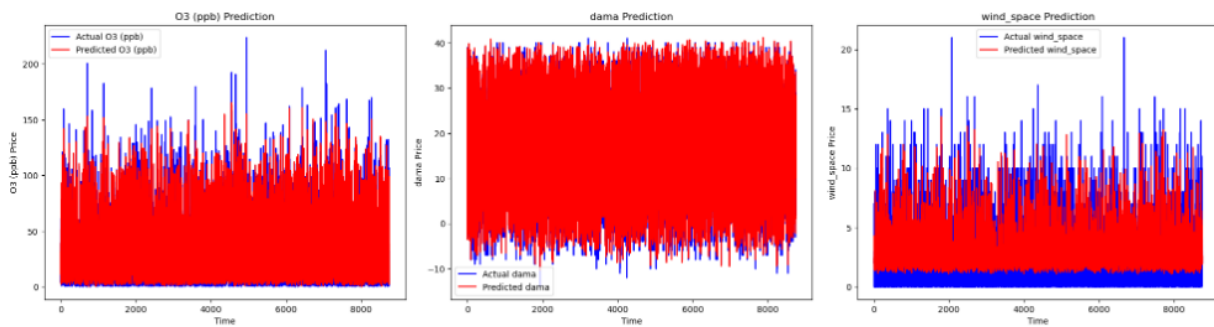
جدول 18- برازش مدل بر روی داده‌های آزمایش

داده‌ها به شکل اولیه باز گردانده شده است و بخشی از آن در زیر مشاهده می‌شود.

```
array([[31.123056 , 34.223145 , 2.4483318],
       [11.035972 , 30.62605 , 1.9833013],
       [39.153606 , 13.470684 , 2.394112 ],
       [ 2.2287505, -3.508111 , 1.8367059],
       [51.38901 , 22.324823 , 3.4888144],
       [19.325123 , 23.400572 , 3.3147278],
       [30.679995 , 22.300163 , 2.61893 ],
       [78.005714 , 38.83319 , 4.281497 ],
       [ 9.139223 , 18.474232 , 2.777791 ],
       [30.57865 , 12.469885 , 2.2080843]], dtype=float32)
```

جدول 19- بازگردانی داده‌ها به شکل قبل از مقیاس کردن

نمودار مقادیر پیش بینی شده در مقابل مقادیر واقعی در شکل زیر آورده شده است. با توجه به نمودار مدل به خوبی برازش داده شده و مقادیر پیش بینی شده با مقادیر واقعی نزدیک به هم هستند.



شکل 21- نمودار مقادیر پیش بینی شده در مقابل مقادیر واقعی در مدل LSTM

برای بررسی کارایی مدل میتوان از معیارهای MAE ، RMSE و  $R^2$  استفاده کرد.

نتیجه گیری:

نتایج تحلیل نشان داد مدل LSTM بهترین برآزش را ارائه می‌دهد، به طوری که قادر بود تغییرات ساعتی آلاینده‌ی ازن را با دقت بالا پیش‌بینی کند. مدل‌های SARIMAX و VAR دقت کمتری نسبت به مدل LSTM ارائه کردند.

با استفاده از مدل LSTM، قابلیت پیش‌بینی تغییرات ساعتی میزان آلاینده‌ی ازن وجود دارد. پیشنهاد می‌شود از این مدل برای تحلیل و پیش‌بینی سایر متغیرهای اقلیمی نیز استفاده شود.

1. Aghelpour, P and M. Nadi, 2018. Assessing the accuracy of SARIMA model in modeling and long-term forecast of average monthly temperature in different climates of Iran. *Climatological Research*, No. 35, pp. 113- 126
2. Ahhashimi, Sh. 2014. Prediction of monthly rainfall in Kirkuk using artificial neural network and time series models. *Journal of engineering and development*, No. 18, pp.129-142.
3. Asakereh, H. 2007. *Climate Change*, University of Zanjan Pub. 1th edition, Zanjan, 246p.
4. Dayal. D., Swain. S., Gautam. A.K., Palmate. S.S., Pandey. A and Mishra. S.K. 2019. Development of ARIMA Model for Monthly Rainfall Forecasting over an Indian River Basin. *World Environmental and Water Resources Congress*.
5. Geweke, J. and Porter-Hudak, S. 1983. The estimation and application of long memory time series models. *Journal of Time Series Analysis*, No. 4, pp. 221–238.
6. Haji Bigloo, M, A. Ghezel Sofloo and H. Ali Mirzaei, 2013. Investigation and forecast of average monthly rainfall using SARIMA technique (Case study: Baba Aman Bojnourd rainfall station). *Engineering and Irrigation Sciences (Scientific Journal of Agriculture)*, No. 3, pp. 41-54.
7. Khosravi, Y, S. Biliyani and A. Bayat, 2017. Temporal analysis of annual rainfall in Shiraz using time series analysis. *Water Resources Engineering*, No, 38, pp. 1-14.