

## پیش‌بینی عمق نوری آئروسول ماهواره‌ای با استفاده از داده‌های پارامترهای اقلیمی

مسعود سلیمانی - دانشجوی دکتری سنجش از دور و GIS، دانشگاه تهران  
میثم ارگانی\* - استادیار گروه سنجش از دور و سیستم اطلاعات جغرافیایی دانشکده جغرافیا، دانشگاه تهران  
رامین پاپی - دانشجوی دکتری سنجش از دور و GIS، دانشگاه تهران  
فاطمه امیری - دانشجوی کارشناسی ارشد سنجش از دور و GIS، دانشگاه تربیت مدرس

دریافت: ۱۳۹۹/۱۱/۱۸ پذیرش: ۱۴۰۰/۰۴/۲۶

### چکیده

عمق نوری آئروسول (AOD) پارامتر سنجش از دور مهمی است که به‌عنوان نماینده‌ای از غلظت آئروسول اتمسفری برای نظارت بر طوفان‌های گردوغبار استفاده می‌شود. در مطالعات پیشین ارتباط بین پارامترهای اقلیمی و AOD گزارش شده است. از طریق تجزیه و تحلیل این ارتباط می‌توان الگوهای مکانی-زمانی AOD را پیش‌بینی کرد. در پژوهش حاضر برای اولین بار از الگوریتم داده‌های M5P نظر به کاربرد آن در خصوص کشف اطلاعات ارزشمند از میان مجموعه داده‌های بزرگ برای استخراج مدل‌های پیش‌بینی‌کننده AOD استفاده شد. بدین منظور، سری زمانی روزانه داده‌های سنجش از دوری پارامترهای دمای هوا، بارش، رطوبت نسبی، و سرعت باد و AOD در یک بازه زمانی ده‌ساله (۲۰۰۵-۲۰۱۴) در محدوده شهرستان اهواز به‌عنوان ورودی‌های M5P تهیه و آماده‌سازی شد. از طریق تشکیل درخت‌های تصمیم مبتنی بر قوانین «اگر-آنگاه» و تجزیه و تحلیل رگرسیون چندمتغیره در چارچوب الگوریتم M5P، چهار مدل پیش‌بینی‌کننده خطی به‌دست آمد. برای اعتبارسنجی مدل‌های خطی، از آماره‌های ضریب همبستگی پیرسون، MAE، و RMSE بهره گرفته شد. مقادیر این آماره‌ها به ترتیب ۰/۶۹، ۰/۲۲، و ۰/۳۱ برآورد شد که حاکی از قابلیت اطمینان مدل‌ها در رابطه با پیش‌بینی AOD است. به‌طور کلی، نتایج این پژوهش نشان داد تکنیک داده‌کاوی در زمینه پیش‌بینی AOD کارآمد است.

واژگان کلیدی: پارامترهای اقلیمی، داده‌کاوی، سنجش از دور، عمق نوری آئروسول، M5P

### مقدمه

ذرات آئروسول تروپوسفری به‌طور مستقیم از طریق پخش و پراکنش و جذب تابش خورشیدی و به‌طور غیرمستقیم با تعدیل خواص میکروفیزیکی و تابشی ابرها نقش مهمی در بیلان انرژی تابشی زمین ایفا می‌کنند. در سطوح پایین تروپوسفر، ذرات آئروسول سبب کاهش کیفیت هوا، کاهش قابلیت دید و خطرهای بهداشت عمومی می‌شوند (لی و همکاران، ۲۰۰۵). به دنبال پیشرفت تکنولوژی، سنجش از دور ماهواره‌ای ابزاری کارآمد برای پایش و اندازه‌گیری الگوهای مکانی-زمانی توزیع آئروسول در مقیاس جهانی فراهم کرده است (کینگ و همکاران، ۱۹۹۲؛ کافمن و همکاران، ۱۹۹۷). عمق نوری آئروسول (AOD) مبتنی بر داده‌های سنجش از دور ماهواره‌ای یک برآورد کمی از مقدار آئروسول موجود در اتمسفر است و می‌توان از آن به‌عنوان شاخصی از غلظت ذرات آئروسول استفاده کرد. به‌طور کلی، AOD میزان انقراض یا خاموشی<sup>۱</sup> یک پرتو نوری هنگام عبور از اتمسفر را اندازه‌گیری می‌کند که با افزایش AOD میزان نرخ خاموش شدن نور نیز افزایش می‌یابد. بر همین اساس، یکی از کاربردهای گسترده AOD در زمینه مطالعات گردوغبار و شناسایی کانون‌های مولد گردوغبار بر اساس تحلیل الگوهای مکانی-زمانی AOD است.

Email: argany@ut.ac.ir

\* نویسنده مسئول، تلفن: ۰۹۱۲۳۲۴۶۰۸۷

1. Aerosol Optical Depth  
2. Extinction of the light

طبق بررسی‌ها، در مطالعات پیشین، مناطق خشک و نیمه‌خشکی که میزان AOD در اتمسفر آن‌ها، در زمان‌های مختلف از سال، بالا بوده است کانون گردوغبار شناسایی کرده‌اند. بدین طریق، برای اولین بار پروسپرو و همکاران (۲۰۰۲) موقعیت کانون‌های مولد گردوغبار جهانی را بر اساس وضعیت آئروسول اتمسفر مبتنی بر به‌کارگیری داده‌های سنجنده<sup>۱</sup> TOMS/ ماهواره<sup>۲</sup> Nimbus 7 شناسایی و نقشه‌برداری کرده‌اند. قابلیت‌های بالای سنجنده‌های MODIS<sup>۳</sup> مستقر بر روی ماهواره‌های Terra و Aqua شامل پوشش محدوده طیفی وسیع، پوشش جهانی، و قدرت تفکیک مکانی بالا در قیاس با سایر مجموعه داده‌های ماهواره‌ای مشابه، اطلاعات کافی به‌منظور بازیابی هم‌زمان ستون آئروسول، بخار آب و ازن اتمسفری را فراهم کرده است. محصولات آئروسول سنجنده‌های MODIS نظیر AOD به‌صورت عملیاتی روی اقیانوس‌ها و خشکی‌ها استخراج می‌شوند و برای مدل‌های گردش عمومی اتمسفر و مطالعات اقلیمی مناسب‌اند (چو و همکاران، ۲۰۰۲). بدین ترتیب، دومین نقشه از موقعیت کانون‌های مولد گردوغبار جهانی با دقت مکانی بالا نسبت به مطالعات گذشته، با استفاده از محصول AOD سنجنده<sup>۴</sup> MODIS را جینوکس و همکاران (۲۰۱۲) تهیه کرده‌اند. به‌طور کلی، مرور منابع پیشین حاکی از اهمیت بالای محصولات آئروسول مبتنی بر داده‌های سنجنش از دور ماهواره‌ای در مدل‌سازی الگوهای مکانی- زمانی گردوغبار و خصوصاً شناسایی کانون‌های مولد گردوغبار است (هسو و همکاران، ۲۰۰۴؛ جینوکس و همکاران، ۲۰۱۰؛ یو و همکاران، ۲۰۱۸؛ رحمتی و همکاران، ۲۰۲۰).

نتایج بررسی‌های انجام‌گرفته نشان می‌دهد سازگاری خوبی بین اندازه‌گیری‌های ماهواره‌ای AOD و اندازه‌گیری‌های زمینی آن توسط شبکه<sup>۵</sup> رباتیک آئروسول (AERONET) وجود دارد (سایر و همکاران، ۲۰۱۳). با وجود این، مزایای بهره‌مندی از AOD ماهواره‌ای برای شناسایی رویدادهای گردوغباری در مناطق خشک با میزان ابرناکی نسبتاً اندک آسمان امکان‌پذیر است (پروسپرو و همکاران، ۲۰۰۲). همچنین، وجود ابر در آسمان اندازه‌گیری‌های زمینی و ماهواره‌ای AOD را با محدودیت جدی مواجه می‌کند. به این ترتیب، مجموعه داده‌های AOD به دلیل عواملی مانند ابرناکی بعضاً دارای شکاف<sup>۶</sup> هستند. از آنجا که امکان پایش و اندازه‌گیری آئروسول در شرایط آسمان ابری با محدودیت مواجه است، استفاده از مجموعه داده‌های نماینده<sup>۷</sup> برای پُر کردن شکاف مزبور به‌عنوان یک مزیت خواهد بود (تان و همکاران، ۲۰۱۶). در این راستا، در مطالعات متعدد مبتنی بر تجزیه و تحلیل داده‌های ماهواره‌ای، بر وجود ارتباط بین پارامترهای اقلیمی و رویدادهای گردوغباری در مناطق مختلف تأکید شده است (ژائو و همکاران، ۲۰۰۴؛ دو و همکاران، ۲۰۰۹). رابطه<sup>۸</sup> بین مقادیر AOD به‌عنوان نمایه‌ای از رخداد گردوغبار و پارامترهای اقلیمی اعم از بارش، دما، فشار، و سرعت باد را نامداری و همکاران (۲۰۱۸) در نواحی بیابانی سوریه، عراق، و عربستان در یک بازه زمانی پانزده‌ساله مطالعه و بررسی کرده‌اند. از طریق تجزیه و تحلیل رگرسیون دومتغیره خطی، همبستگی قابل‌قبولی بین دما و بارش و AOD مشاهده شده است. بنابراین، با در نظر گرفتن ارتباط بین پارامترهای اقلیمی و AOD، می‌توان از این پارامترها به‌عنوان یک مجموعه داده<sup>۹</sup> نماینده برای تخمین مقادیر AOD برای مناطق فاقد داده یا دارای ابر استفاده کرد. افزون بر این، نظر به صحت و کارایی مطلوب اندازه‌گیری‌های ماهواره‌ای AOD برای پایش طوفان‌های گردوغبار، پیش‌بینی الگوهای مکانی- زمانی AOD برای اتخاذ تدابیر مدیریت بحران مناسب و به‌موقع در راستای کنترل یا کاهش ابعاد زیست‌محیطی مخرب و زیان‌بار آن حائز اهمیت است. به این صورت که با استفاده از پیش‌بینی مقادیر پارامترهای اقلیمی و تجزیه و تحلیل روابط

1. Total Ozone Mapping Spectrometer
2. Moderate Resolution Imaging Spectroradiometer
3. Aerosol Robotic Network
4. Gap
5. Proxy

می‌توان مقادیر AOD را پیش‌بینی کرد. بر این اساس، برای دستیابی به نتایج پیش‌بینی قابل‌اعتماد AOD، بهره‌گیری از یک رویکرد مدل‌سازی تعمیم‌پذیر، که بتواند علاوه بر مدل‌کردن روابط بعضاً پیچیده بین مجموعه‌عظیم داده‌های ورودی، به نحو رضایت‌بخشی معضلات مشروح را مرتفع سازد، ضرورت دارد.

داده‌کاوی<sup>۱</sup> به‌عنوان فرایند کشف دانش در پایگاه‌داده (KDD)<sup>۲</sup> تلفیقی از تکنولوژی‌های پایگاه‌داده، تجزیه‌وتحلیل‌های آماری، و یادگیری ماشین است. از جمله عملکردهای اصلی داده‌کاوی می‌توان به خلاصه‌سازی<sup>۳</sup>، قوانین انجمنی<sup>۴</sup>، طبقه‌بندی<sup>۵</sup>، پیش‌بینی<sup>۶</sup>، و خوشه‌بندی<sup>۷</sup> اشاره کرد (ژائو، ۲۰۰۳). در چند سال اخیر رویکردهای مختلف مبتنی بر داده‌کاوی در زمینه مخاطرات محیطی (گسوامی و همکاران، ۲۰۱۸) اعم از پیش‌بینی آلاینده‌های جوی (سیوک و اوسوسکی، ۲۰۱۶؛ بلینگر و همکاران، ۲۰۱۷)، نقشه‌برداری کانون‌های مستعد گردوغبار (غلامی و همکاران، ۲۰۲۰)، نقشه‌برداری مناطق مستعد زمین‌لغزش (لی و همکاران، ۲۰۱۷)، و فرونشست زمین (قربان‌زاده و همکاران، ۲۰۱۸) به‌کار گرفته شده است. به‌طور کلی، در مطالعات نام‌برده بر توانمندی و کاربرد گسترده مدل‌های داده‌کاوی در زمینه استخراج اطلاعات بازارش بر مبنای تجزیه‌وتحلیل مجموعه‌داده‌های بزرگ تأکید شده است.

در این مطالعه، برای تجزیه‌وتحلیل و استخراج روابط بین پارامترهای اقلیمی و AOD، به‌منظور به‌دست‌آوردن مدل‌های پیش‌بینی‌کننده، یکی از الگوریتم‌های داده‌کاوی تحت عنوان MSP مدنظر قرار گرفت. الگوریتم MSP ترکیبی از مدل‌های درختی و رگرسیونی است. به‌طور کلی، این الگوریتم متشکل از یک ساختار درختی با مدل‌های رگرسیون خطی در برگ‌هاست. این مدل‌های رگرسیون در واقع معادلات رگرسیون چندمتغیره خطی‌اند که می‌توانند برای یافتن مقادیر پیش‌بینی‌شده حل شوند. ساختارهای درختی معمولاً بسیار بزرگ‌تر و پیچیده‌تر از معادلات رگرسیونی‌اند، اما در عوض صحت پیش‌بینی به مراتب بالاتری را ارائه می‌دهند. در مقابل تفسیر نتایج مدل‌های درختی به دلیل اندازه زیاد آن‌ها معمولاً دشوار است. از این رو، در الگوریتم MSP تلفیقی از مدل‌های درختی و رگرسیونی برای برخورداری هم‌زمان از صحت پیش‌بینی بالا و سهولت در تفسیر نتایج استفاده شده است (فرانک و همکاران، ۱۹۹۸؛ شبان و همکاران، ۲۰۱۶).

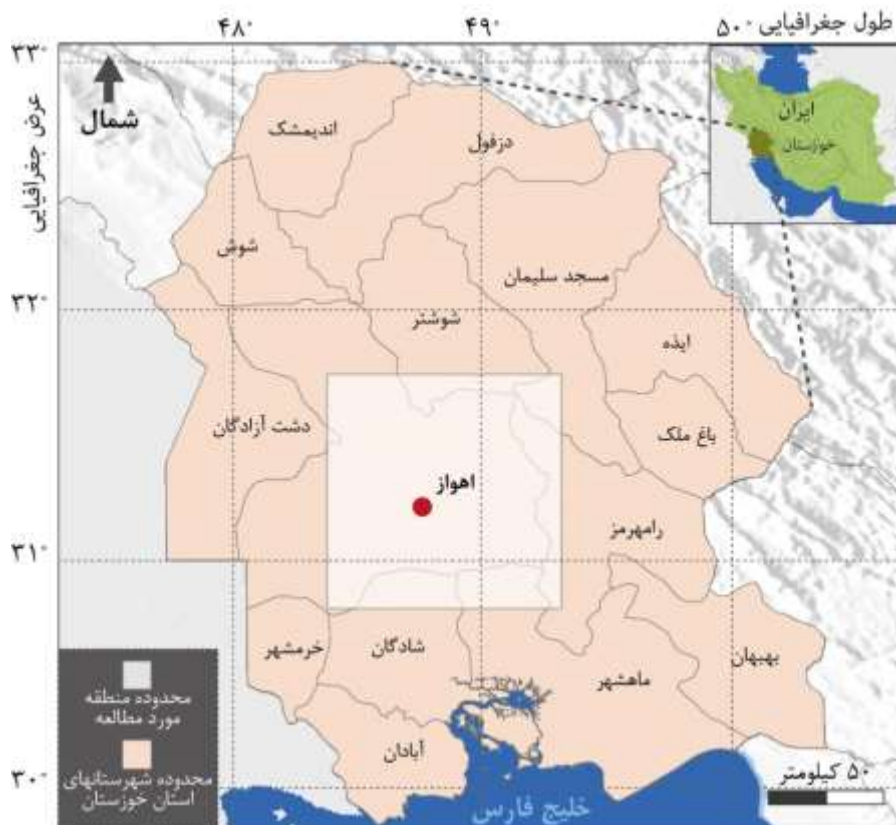
پژوهش حاضر با هدف ارائه مدل‌های پیش‌بینی‌کننده AOD بر اساس تجزیه‌وتحلیل ارتباط بلندمدت بین مقادیر پارامترهای اقلیمی شامل دما، بارش، رطوبت نسبی، و سرعت باد با مقدار AOD سامان‌دهی شده است. استان خوزستان به‌عنوان مطالعه موردی و یکی از مناطق بسیار متأثر از طوفان‌های گردوغبار در ایران مدنظر قرار گرفت. برای نیل به این هدف، از سری زمانی روزانه داده‌های سنجش از دور اعم از محصول AOD سنجنده MODIS و داده‌های اقلیمی در فاصله زمانی سال‌های ۲۰۰۵ تا انتهای ۲۰۱۴ استفاده شد. بدین منظور، الگوریتم داده‌کاوی MSP نظر به قابلیت‌های ویژه آن در زمینه پیش‌بینی و طبقه‌بندی مدنظر قرار گرفت.

1. Data mining
2. knowledge Discovery in Databases
3. Summarization
4. Association rules
5. Classification
6. Prediction
7. Clustering

## مواد و روش‌ها

### منطقه مورد مطالعه

منطقه مورد مطالعه در این پژوهش استان خوزستان (بخش‌هایی از شهرستان‌های اهواز، شادگان، ماهشهر، رامهرمز، و شوشتر) واقع در جنوب غربی ایران است (محدوده خاکستری رنگ در شکل ۱). محدوده انتخابی به کانون‌های درون مرزی و فرامرزی گردوغبار نزدیک است. از طرفی، این محدوده در معرض چندین سیستم گردوغبار قرار دارد و هر ساله طوفان‌های گردوغبار زیادی در آنجا رخ می‌دهد. در برخی شهرستان‌های واقع در محدوده مورد مطالعه، نظیر اهواز، بیش از ۳۱ روز از سال گردوغبار رخ می‌دهد. این مهم عده‌ای از مشکلات اقتصادی-اجتماعی و زیست‌محیطی را در سالیان گذشته برای ساکنان مناطق نام‌برده همراه داشته است (نجفی و همکاران، ۲۰۱۴؛ صمدی و همکاران، ۲۰۱۴).



شکل ۱. منطقه مورد مطالعه: استان خوزستان واقع در جنوب غربی ایران - محدوده خاکستری رنگ با ابعاد تقریبی ۱۰۰ × ۱۰۰ کیلومتر برای تجزیه و تحلیل و مدل‌سازی رفتار بلندمدت پارامترهای اقلیمی و AOD مدنظر قرار گرفته است

### داده‌های مورد استفاده

به منظور تجزیه و تحلیل الگوهای زمانی رفتار پارامترهای اقلیمی و AOD از محصولات مبتنی بر داده‌های سنجش از دور بهره گرفته شد. مشخصات داده‌های مورد استفاده در جدول ۱ ارائه شده است. برای افزایش تعمیم‌پذیری نتایج مدل‌سازی از طریق لحاظ کردن شرایط محیطی متفاوت، دوره زمانی مورد بررسی به گونه‌ای تعیین شد که هم سال‌های پُرگردوغبار را پوشش دهد و همچنین داده‌های سنجش از دور روزانه در دسترس باشد. در دو دهه اخیر ناشی از رویدادهای دوره‌ای خشک‌سالی و اقدامات انسانی در کشورهای منطقه غرب آسیا طوفان‌های گردوغبار شدیدی رخ داده است. به عنوان نمونه

بارز آن می‌توان به خشک‌سالی شدید اتفاق افتاده در فاصله زمانی سال‌های ۲۰۰۷ تا ۲۰۰۹ در این منطقه اشاره کرد (تریگو و همکاران، ۲۰۱۰). در بازه زمانی مزبور، خصوصاً سال ۲۰۰۸، فعالیت‌های گردوغباری نیز در کانون‌های منطقه غرب آسیا افزایش چشمگیری داشته است. قبل و بعد از این دوره خشک‌سالی، فعالیت‌های گردوغبار با شدت و فراوانی کمتری اتفاق افتاده است (نبوی و همکاران، ۲۰۱۶). بر این اساس، یک بازه زمانی ده‌ساله از ۲۰۰۵ تا ۲۰۱۴ (معادل ۱۳۸۴ تا ۱۳۹۳) مدنظر قرار گرفت. به این صورت که از داده‌های ۲۰۰۵/۱/۱ تا ۲۰۱۳/۱۲/۳۰ (۹۰ درصد داده‌ها) برای مدل‌سازی و از ۲۰۱۴/۱/۱ تا ۲۰۱۴/۱۲/۳۰ (۱۰ درصد داده‌ها) برای اعتبارسنجی استفاده شد. بدین ترتیب، برای چهار پارامتر اقلیمی شامل بارش، دمای سطح زمین، رطوبت نسبی، و سرعت باد تعداد ۱۴۶۰۰ تصویر و ۳۶۵۰ تصویر نیز برای AOD تهیه شد. در ادامه سری زمانی تصاویر به اندازه محدود مورد مطالعه (شکل ۱) برش داده شد. سپس، مقادیر پیکسل‌های تصاویر سری زمانی مربوط به هر یک از پارامترها برای تجزیه و تحلیل آماری استخراج شد. مطابق با مندرجات جدول ۱، برای داده‌هایی با قدرت تفکیک مکانی بالاتر اعم از بارش، سرعت باد، و AOD، تعداد پیکسل‌های هر تصویر در محدوده مورد مطالعه به ترتیب برابر با ۱۶، ۱۶، و ۱۰۰۰۰ بود که از آن‌ها به تفکیک هر پارامتر میانگین گرفته شد.

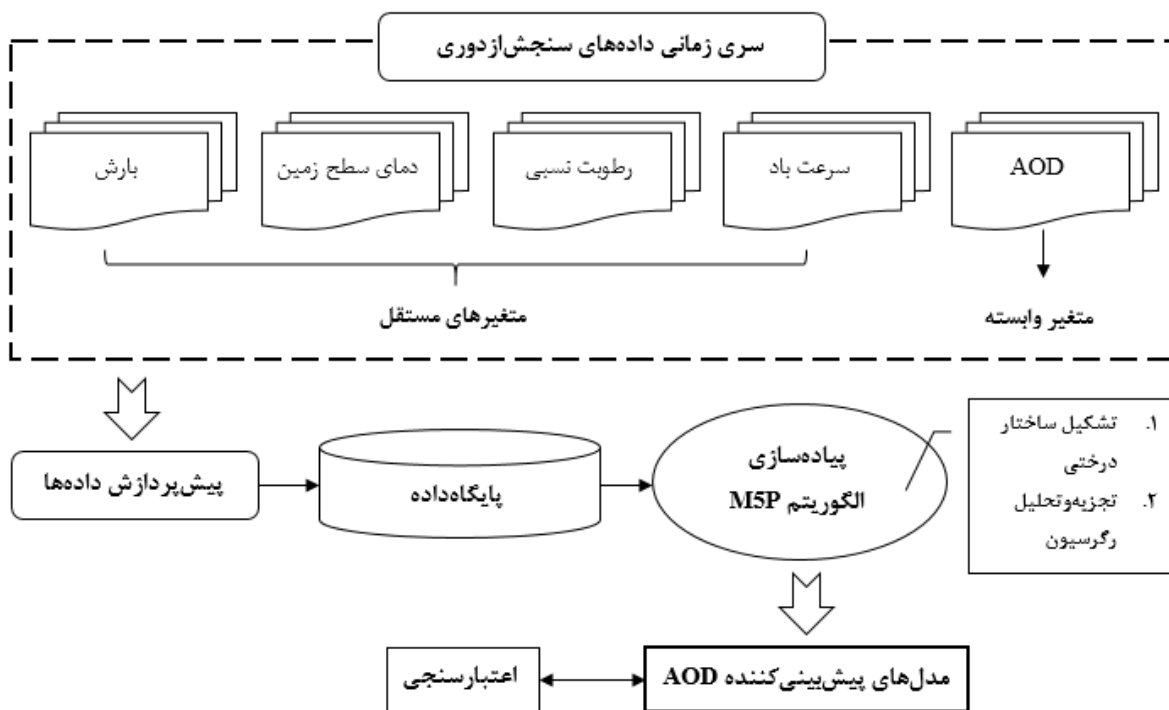
جدول ۱. مشخصات داده‌های سنجش از دور مورد استفاده در بازه زمانی ۲۰۰۵ تا ۲۰۱۴

پارامتر	مشخصه	واحد	قدرت تفکیک مکانی	قدرت تفکیک زمانی	منبع
بارش	3B42 v7	میلی‌متر بر ساعت	۰/۲۵ درجه	۳ ساعته	TRMM
دمای سطح زمین	AIRX3STD v006	سانتی‌گراد	۱ درجه	روزانه	AIRS
رطوبت نسبی	AIRX3STD v006	درصد	۱ درجه	روزانه	AIRS
سرعت باد	CLSM025_D v2.0	متر بر ثانیه	۰/۲۵ درجه	روزانه	GLDAS
عمق نوری آتروسول	MCD19A2	-	۱ کیلومتر	روزانه	MODIS

## روش تحقیق

در این مطالعه به منظور استخراج مدل‌های پیش‌بینی‌کننده AOD بر اساس پارامترهای اقلیمی، از الگوریتم داده‌کاوی M5P مبتنی بر ساختار درختی و تجزیه و تحلیل رگرسیون خطی چندمتغیره استفاده شد. بر این اساس، یک پایگاه داده مکانی متشکل از داده‌های سری زمانی سنجش از دور مربوط به چهار پارامتر اقلیمی (به‌عنوان متغیرهای مستقل) شامل دمای هوای سطح زمین (SAT)<sup>۱</sup>، بارش (P)<sup>۲</sup>، رطوبت نسبی سطح زمین (SRH)<sup>۳</sup>، و سرعت باد (WS)<sup>۴</sup> و AOD به‌عنوان متغیر وابسته تولید شد. برای اجرای مدل M5P از نرم‌افزار وکا (WEKA)<sup>۵</sup> استفاده شد. پس از تجزیه و تحلیل روابط بین متغیرهای مستقل و وابسته از طریق ساختار مدل درختی و رگرسیون چندمتغیره خطی، قوانین پیش‌بینی‌کننده AOD استخراج شد. برای اعتبارسنجی مدل‌های پیش‌بینی‌کننده خطی از شاخص‌های آماری استفاده شد. ساختار کلی پژوهش در شکل ۲ ارائه شده است.

1. Surface Air Temperature
2. Precipitation
3. Surface Relative Humidity
4. Wind Speed
5. Waikato Environment for Knowledge Analysis



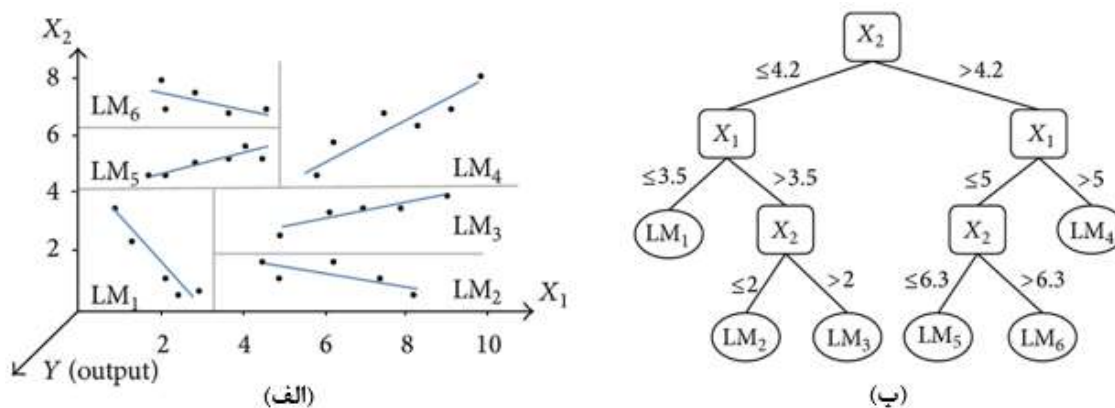
شکل ۲. فلوجارت روند اجرای پژوهش

### الگوریتم MSP

به‌طور کلی، مدل‌های مبتنی بر ساختار درختی بر اساس روش تقسیم و غلبه<sup>۱</sup> برای ایجاد روابط بین متغیرهای مستقل و وابسته ساخته می‌شوند (پال، ۲۰۰۶). مدل درختی MSP جزو رویکردهای یادگیری استقرایی<sup>۲</sup> به‌شمار می‌رود. این رویکردها از طریق تشکیل قوانین اگر-آنگاه<sup>۳</sup> گروه‌های خواص منحصر به فرد به اشتراک گذاشته شده توسط مجموعه نمونه‌ها را شناسایی می‌نمایند. رویکردهای یادگیری استقرایی برای کشف روابط بین داده‌ها و به‌دست آوردن دانش ساختاری از آن‌ها کارآمدند. مدل MSP در وهله<sup>۴</sup> نخست به منظور تولید دانش دامنه<sup>۴</sup> یک درخت تصمیم می‌سازد. این مدل به‌طور کلی ترکیبی از قابلیت‌های درختان تصمیم متعارف و رگرسیون خطی<sup>۵</sup> است. توابع خطی در گره‌های درخت استفاده می‌شود. به جای به حداکثر رساندن اطلاعات کسب‌شده، در هر گره از یک معیار تقسیم استفاده می‌شود که تغییر درون-زیرمجموعه<sup>۶</sup> در مقادیر کلاس در هر شاخه را به حداقل می‌رساند (آندینا و فام، ۲۰۰۷). وهله<sup>۵</sup> دوم، پس از ساختن درخت، هرس کردن<sup>۷</sup> است و به دنبال آن روش نرم‌کنندگی برای جلوگیری از ناپیوستگی‌های شدید انجام می‌شود. درخت تصمیم مدل MSP طبقه‌بندی‌کننده‌های<sup>۸</sup> دقیقی ایجاد می‌کند، به‌ویژه هنگامی که بسیاری از خصوصیات عددی باشند

1. divide-and-conquer
2. Inductive Learning Method
3. If - then
4. Domain Knowledge
5. Linear Regression
6. Intra-Subset
7. Pruning
8. Classifier

(اپرا و همکاران، ۲۰۱۶). در واقع، در M5P معادلات رگرسیون خطی چندمتغیره به ازای زیرمجموعه‌های مختلفی از متغیرهای مستقل ورودی حل می‌شوند. شکل ۳- الف تقسیم فضای ویژگی حاصل از برآزش دو متغیر مستقل  $x_1$  و  $x_2$  به شش مدل خطی (LM) رگرسیونی بر اساس الگوریتم M5P را نشان می‌دهد. شکل ۳- ب نحوه تقسیم فضای متغیرهای مستقل ورودی و روابط شاخه‌ها و برگ‌ها را در ساختار درختی مدل M5P نمایش می‌دهد (رحیمی خوب و همکاران، ۲۰۱۳).



شکل ۲. مفهوم شماتیکی مدل درختی M5P برای یافتن قوانین پیش‌بینی کننده بر اساس تجزیه و تحلیل رگرسیون خطی چندمتغیره از طریق تقسیم فضای متغیرهای مستقل ورودی با استفاده از تشکیل سلسله قوانین اگر- آنگاه (رحیمی خوب و همکاران، ۲۰۱۳)

## اعتبارسنجی

در مطالعه حاضر برای ارزیابی ارتباط بین متغیرهای ورودی مدل M5P و همچنین قابلیت پیش‌بینی مدل، از سه شاخص آماری شامل ضریب همبستگی پیرسون ( $r$ )<sup>۲</sup>، میانگین خطای مطلق (MAE)<sup>۳</sup>، و خطای جذر میانگین مربعات (RMSE)<sup>۴</sup> استفاده شد. آماره  $r$  برای اندازه‌گیری ارتباط خطی بین مقادیر متغیرهای ورودی (پارامترهای اقلیمی و AOD) به کار گرفته شد. مقدار این آماره بین -۱ تا +۱ متغیر است. مقادیر مثبت بیانگر رابطه مستقیم و مقادیر منفی به معنی رابطه معکوس است. مقدار +۱ بیانگر همبستگی مستقیم و کامل، -۱ بیانگر همبستگی معکوس و کامل، و مقدار صفر بیانگر عدم همبستگی و استقلال کامل متغیرها از همدیگر است (شوبر و همکاران، ۲۰۱۸). فرض کنید  $X$  و  $Y$  دو مجموعه از نمونه‌ها هستند که هر یک از آن‌ها شامل  $n$  تعداد مشاهده به صورت  $X = (x_1, x_2, \dots, x_n)$  و  $Y = (y_1, y_2, \dots, y_n)$  هستند. بر این اساس، ضریب همبستگی پیرسون به صورت رابطه ۱ تعریف می‌شود (خو و دنگ، ۲۰۱۷):

$$r = \frac{(N \sum x_i y_i - \sum x_i \sum y_i)}{\sqrt{N x_i^2 - (\sum x_i)^2} \sqrt{N y_i^2 - (\sum y_i)^2}} \quad \text{رابطه ۱}$$

1. Linear Model
2. Pearson Correlation Coefficient
3. Mean Absolute Error
4. Root Mean Square Error

در خصوص مدل‌های درختی، مقادیر پیش‌بینی‌شده در نمونه‌های آزمون<sup>۱</sup> که در تکرارهای مختلف مدل از طریق رویکرد اعتبارسنجی متقاطع<sup>۲</sup> (تعدادی تکرار<sup>۳</sup> برای هرس کردن) به‌دست می‌آید عبارت‌اند از:  $P_1, P_2, \dots, P_n$  و مقادیر واقعی عبارت‌اند از:  $a_1, a_2, \dots, a_n$ . بدین ترتیب،  $P_i$  بیانگر مقدار عددی پیش‌بینی برای نمونه آزمون  $i$ ام است. بر این اساس، می‌توان آماره MAE را تعریف کرد. MAE میانگین خطاها را در مجموعه‌ای از پیش‌بینی‌ها اندازه‌گیری می‌کند و بدین ترتیب یک امتیاز خطی<sup>۴</sup> ایجاد می‌کند. این امتیاز به این معنی است که همه تفاوت‌های منحصر به فرد به‌طور متوسط در میانگین وزن‌دهی می‌شوند. این آماره عبارت است از میانگین بزرگی خطاهای منفرد بدون در نظر گرفتن علامت آن‌ها که از طریق رابطه<sup>۲</sup> محاسبه می‌شود (ویتن و فرانک، ۲۰۰۲):

$$MAE = \frac{|p_1 - a_1| + \dots + |p_n - a_n|}{n} \quad \text{رابطه ۲}$$

RMSE میانگین خطاها را اندازه‌گیری می‌کند. این آماره وزن نسبتاً بالایی به خطاهای بزرگ می‌دهد که از طریق رابطه<sup>۳</sup> به‌دست می‌آید:

$$RMSE = \sqrt{\frac{(p_1 - a_1)^2 + \dots + (p_n - a_n)^2}{n}} \quad \text{رابطه ۳}$$

که در رابطه‌های ۲ و ۳  $P$  بیانگر مقادیر پیش‌بینی‌شده،  $a$  مقادیر واقعی،  $\bar{a}$  مقادیر میانگین واقعی، و  $n$  تعداد متغیرهاست. به‌طور کلی، برای تشخیص تغییرات خطا در مجموعه پیش‌بینی‌ها می‌توان از آماره‌های MAE و RMSE استفاده کرد. دامنه عددی مقادیر این دو آماره می‌تواند بین صفر تا بی‌نهایت متغیر باشد (ویلموت و ماتسورا، ۲۰۰۵).

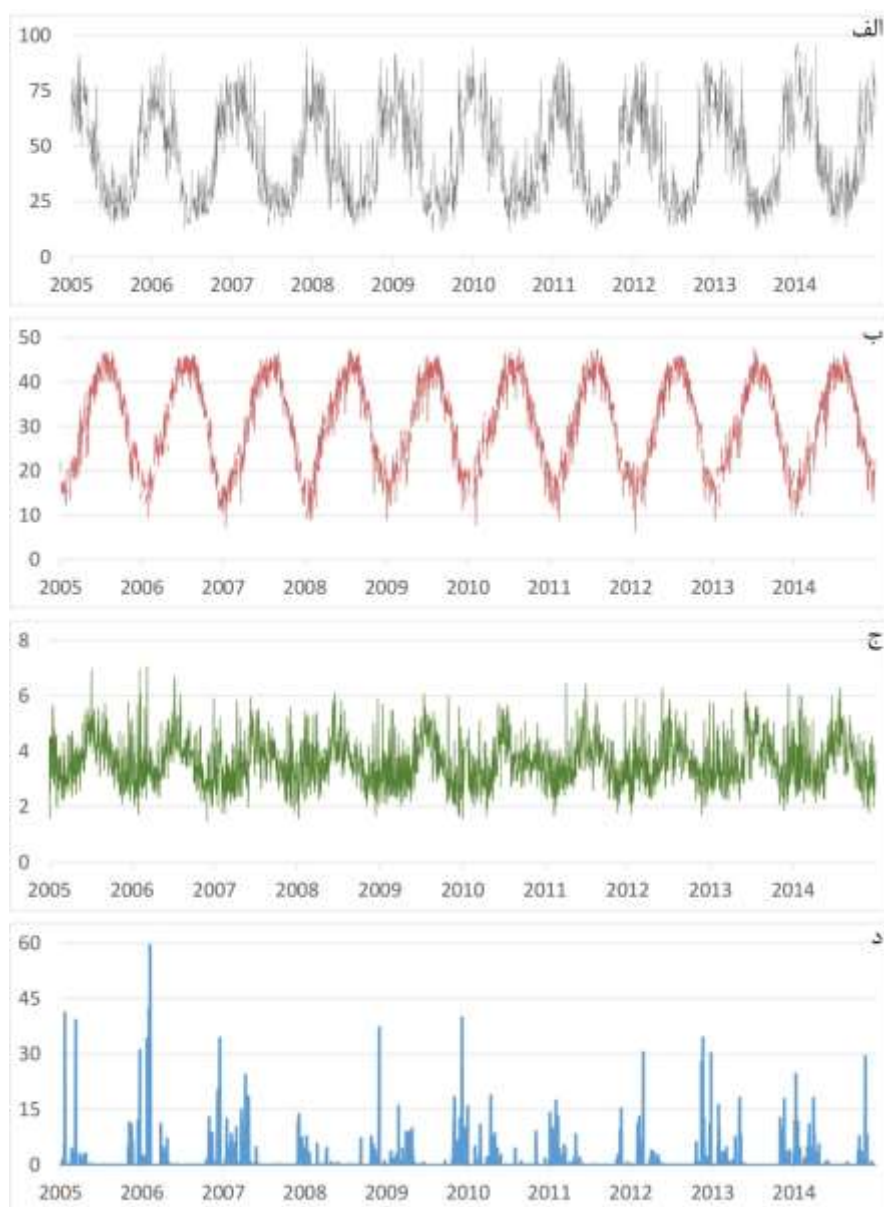
## بحث و نتایج

### استخراج مدل‌های پیش‌بینی‌کننده

در این مطالعه ارائه مدل‌های پیش‌بینی‌کننده AOD به‌عنوان شاخصی از وضعیت گردوغبار اتمسفر بر اساس تجزیه و تحلیل رفتار بلندمدت پارامترهای اقلیمی و ارتباط آن‌ها با AOD مورد نظر است. پس از پیش‌پردازش سری زمانی داده‌های مربوط به پارامترهای اقلیمی و AOD به‌عنوان مجموعه داده‌های آموزشی، متغیرهای مستقل و وابسته ورودی مدل طبقه‌بندی‌کننده درختی M5P تعریف شد. متغیرهای مستقل ورودی مدل شامل سری زمانی روزانه مقادیر پارامترهای دما، بارش، رطوبت نسبی، و سرعت باد در یک بازه زمانی ده‌ساله در شکل ۴ ارائه شده است. در همه نمودارها روند کلی تقریباً ثابتی وجود دارد. در این میان، صرف‌نظر از روندهای فصلی، برخی نوسانات و روندهای کاهشی و افزایشی در مقادیر پارامترها در برخی برهه‌های زمانی مشهود است. با در نظر داشتن ارتباط قابل توجه بین AOD و مقادیر پارامترهای اقلیمی، این تغییرات و واریانس موجود در داده‌های سری زمانی حاوی اطلاعات ارزشمندی است. بر همین اساس، در پژوهش حاضر برای کشف ارتباطات موجود بین داده‌های سری زمانی و دستیابی به مدل‌های پیش‌بینی‌کننده از قابلیت‌های منحصر به فرد الگوریتم M5P استفاده شد.

1. Test Instances
2. Cross-Validation
3. Fold
4. Linear Score



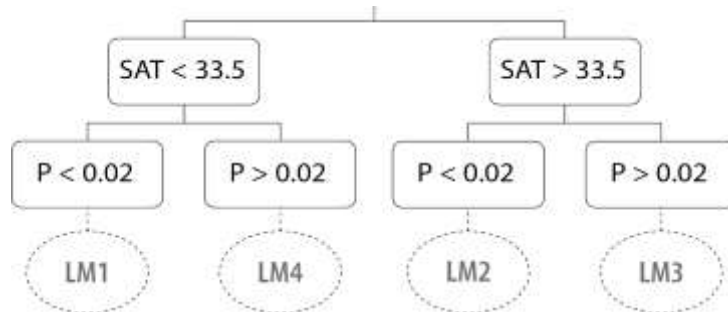


شکل ۴. نمودار سری زمانی روزانه پارامترهای مستقل ورودی مدل MSP در بازه زمانی ۲۰۰۵ تا ۲۰۱۴. الف) رطوبت نسبی، ب) دمای هوای سطح زمین (درصد)؛ ج) دمای هوای سطح زمین (درجه سانتی‌گراد)؛ د) بارش (میلی‌متر)

در ادامه مراحل اجرای الگوریتم MSP اعم از همگن‌سازی مجموعه داده‌های مستقل ورودی از طریق تشکیل درختان تصمیم‌گیری بر اساس سلسله قوانین اگر-آنگاه، تجزیه و تحلیل رگرسیون خطی چندمتغیره در کلاس‌های همگن، و نهایتاً اعتبارسنجی نتایج مدل در محیط نرم‌افزار وکا انجام شد. به منظور حداقل‌سازی انحرافات و بهبود فرایند پردازش، برای دستیابی به بهترین مدل طبقه‌بندی‌کننده / پیش‌بینی‌کننده متناسب با مجموعه داده‌های آموزشی (آلسانتانی، ۲۰۲۰)، درختان تصمیم با ده بار اعتبارسنجی متقابل<sup>۱</sup> اجرا شد (سرینیواسان و مکالا، ۲۰۱۴). درخت تصمیم تولیدشده با استفاده از MSP برای ارائه مدل‌های پیش‌بینی‌کننده AOD در شکل ۵ ارائه شده است. بدین ترتیب، در مجموع، چهار

1. 10 folds-cross validation

مدل خطی یا قانون پیش‌بینی‌کننده<sup>۱</sup> برای تخمین AOD بر اساس مقادیر پارامترهای اقلیمی استخراج شد (جدول ۲). درنهایت، با قراردادن مقادیر پارامترهای اقلیمی در مدل‌های خطی به‌دست‌آمده، می‌توان مقدار AOD را بر اساس آستانه‌های تعریف‌شده توسط الگوریتم M5P برآورد کرد. مدل‌های خطی به‌دست‌آمده قادر به پیش‌بینی مقادیر AOD در شرایط مختلف (بر اساس پارامترهای اقلیمی) است. غالباً طوفان‌های گردوغبار در دماهای بالا و حداقل میزان بارش رخ می‌دهد. بر همین اساس، برای شناسایی طوفان‌های گردوغبار بر مبنای AOD، مدل خطی شماره ۲ (LM-2) از اهمیت بالاتری در قیاس با سایر مدل‌ها برخوردار است.



شکل ۵. ساختار درخت تصمیم M5P تشکیل‌شده برای پیش‌بینی مقادیر AOD بر اساس پارامترهای اقلیمی

جدول ۲. مدل‌های خطی (LM) پیش‌بینی‌کننده AOD بر اساس پارامترهای اقلیمی

شماره قانون	قانون / مدل پیش‌بینی‌کننده خطی
LM-1	$AOD = -0.0064 \times SRH + 0.0156 \times SAT + 0.0612 \times WS - 0.031 \times P + 1.0029$
LM-2	$AOD = -0.0062 \times SRH + 0.0121 \times SAT + 0.1589 \times WS + 0.0045 \times P + 0.779$
LM-3	$AOD = -0.0368 \times SAT + 3.6843$
LM-4	$AOD = 0.1472 \times P + 1.633$

### اعتبارسنجی قوانین پیش‌بینی‌کننده

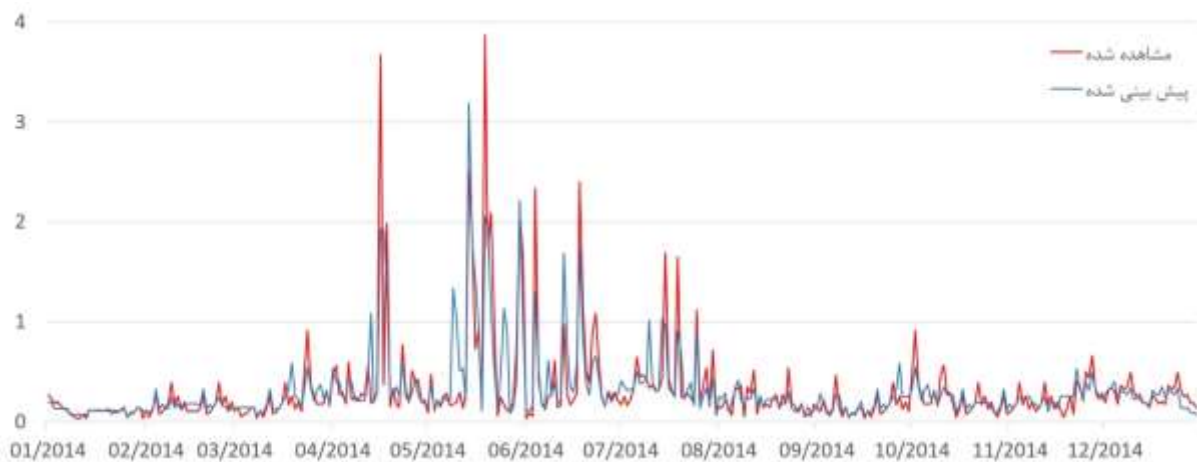
موضوع مهم در خصوص مدل‌های پیش‌بینی‌کننده ارزیابی و تعیین درجه اطمینان آن‌هاست. در این راستا، به‌منظور ارزیابی صحت پیش‌بینی و به‌طور کلی عملکرد الگوریتم M5P در رابطه با پیش‌بینی AOD بر اساس مدل‌های خطی مستخرج، با به‌کارگیری روابط ۱ تا ۳، شاخص‌های آماری  $r$ ، MAE، و RMSE به‌ترتیب برابر با ۰/۶۹، ۰/۲۲، و ۰/۳۱ محاسبه شد. به‌طور کلی، وجود همبستگی بالاتر بین هر یک از پارامترهای اقلیمی با AOD میزان خطای پیش‌بینی مدل‌های خطی استخراج‌شده از طریق درخت تصمیم M5P را کاهش می‌دهد؛ به‌طوری‌که به ازای مقادیر همبستگی بالاتر، مقادیر آماره‌های MAE، و RMSE کاهش خواهد یافت و بالعکس. می‌توان گفت میزان قابلیت اطمینان و صحت مدل‌های خطی ارائه‌شده به میزان همبستگی بین متغیرهای مستقل و وابسته ورودی مدل بستگی دارد. جدول ۳ یک مقیاس پنج‌سطحی برای اندازه‌گیری و توصیف میزان قدرت همبستگی را نشان می‌دهد (آلسانتانی، ۲۰۲۰). بر اساس این تقسیم‌بندی، میزان همبستگی محاسبه‌شده بین پارامترهای اقلیمی و AOD در محدوده همبستگی قوی است. اعتبارسنجی نتایج الگوریتم M5P بر اساس ارزیابی میزان همبستگی و خطاهای پیش‌بینی از طریق آماره‌های MAE و RMSE نشان‌دهنده عملکرد و صحت قابل قبول مدل‌های خطی در رابطه با پیش‌بینی AOD است.

با توجه به پویایی ذرات آئروسول (به‌ویژه گردوغبار) و قابلیت حمل از طریق جریان هوا و باد تا فواصل بسیار دوردست از منبع انتشار آن‌ها (گودی، ۲۰۰۹)، این احتمال وجود دارد که مقدار AOD ثبت‌شده برای یک پیکسل توسط سنسور ماهواره‌ای دقیقاً مربوط به همان منطقه بر روی زمین نباشد (درویشی بلورانی و همکاران، ۲۰۲۱). بنابراین، در رابطه با میزان خطای پیش‌بینی مدل‌ها ذکر این نکته لازم است که این مهم می‌تواند ناشی از قابلیت حمل ذرات آئروسول توسط باد باشد. نظر به ارتباط قوی بین AOD و پارامترهای اقلیمی، ناهمخوانی‌های احتمالی می‌تواند ناشی از دلیل مشروح باشد، زیرا فرضاً طوفان گردوغبار برخاسته از یک کانون ممکن است هیچ ارتباطی با مقادیر پارامترهای اقلیمی در مکان مقصد نداشته باشد.

جدول ۳. تفسیر میزان قدرت همبستگی

تفسیر	مقدار همبستگی
همبستگی بسیار قوی	$\geq \pm 0.7$
همبستگی قوی	$\leq \pm 0.7$ تا $\pm 0.4$
همبستگی متوسط	$\leq \pm 0.4$ تا $\pm 0.3$
همبستگی ضعیف	$\leq \pm 0.3$ تا $\pm 0.2$
فقدان همبستگی	$\leq \pm 0.2$ تا $\pm 0.1$

درنهایت، برای نمایش نتایج اعتبارسنجی و ارزیابی عملکرد مدل‌های خطی به‌صورت بصری از مقادیر پارامترهای اقلیمی موردبررسی از ۲۰۱۴/۱/۱ تا ۲۰۱۴/۱۲/۳۰ به‌عنوان ورودی مدل‌های پیش‌بینی‌کننده خطی (جدول ۲) استفاده شد. مطابق با شکل ۶، نتایج ارزیابی صحت صورت‌گرفته درمجموع حاکی از همگرایی نتایج مدل‌های خطی به مقادیر واقعی AOD در یک بازه زمانی یک‌ساله در منطقه مورد مطالعه است. به‌طور کلی، این همگرایی متضمن قابلیت اطمینان مدل‌های خطی مستخرج از الگوریتم MSP برای پیش‌بینی AOD بر اساس پارامترهای اقلیمی است. بر اساس شکل ۶، میزان اختلاف بین مقادیر پیش‌بینی‌شده و مشاهده‌شده تقریباً از اواخر ماه چهارم سال (آوریل) تا ماه هفتم (جولای) افزایش یافته است. با توجه به وجود کانون‌های فعال گردوغبار در منطقه بین‌النهرین و قابلیت حمل ذرات گردوغبار از طریق وزش بادهای شمال با راستای شمال غرب به جنوب شرق در عراق در فصول بهار و تابستان (درویشی بلورانی و همکاران، ۲۰۲۱)، این مهم باعث افزایش غلظت ذرات گردوغبار در اتمسفر شده و متعاقباً افزایش و تشدید AOD در اتمسفر منطقه مورد مطالعه را سبب شده است. بالاتر بودن مقادیر AOD مشاهداتی در ماه‌های مزبور نسبت به مقادیر پیش‌بینی‌شده بر دلایل مشروح صحنه می‌گذارد.



شکل ۶. ارزیابی عملکرد مدل‌های خطی مستخرج از الگوریتم M5P در رابطه با پیش‌بینی AOD بر اساس مقادیر پارامترهای اقلیمی دما، بارش، رطوبت نسبی، و سرعت باد در محدوده استان خوزستان (محور افقی نمودار بیانگر روز از ماه و محور عمودی نشان‌دهنده میزان AOD است)

### نتیجه‌گیری

به‌طور کلی، در این مطالعه قابلیت الگوریتم داده‌کاوی M5P در زمینه پیش‌بینی AOD ارزیابی شد. با استفاده از الگوریتم M5P مبتنی بر یادگیری استقرایی و به‌کارگیری داده‌های سری زمانی سنجش از دور از طریق تشکیل درختان تصمیم‌گیری بر اساس مجموعه قوانین اگر- آنگاه، چهار مدل پیش‌بینی‌کننده خطی بر مبنای پارامترهای اقلیمی استخراج شد. مدل‌های پیش‌بینی‌کننده با استفاده از داده‌های مربوط به محدوده شهرستان اهواز استخراج و اعتبارسنجی شد.

AOD به‌عنوان شاخصی از وضعیت آئروسول اتمسفر اهمیت بالایی برای انجام‌دادن مطالعات در زمینه طوفان‌های گردوغبار یافته است. دسترسی به داده‌های AOD در برخی مناطق کره زمین و در برخی فصول از سال ناشی از دلایلی من جمله پوشش ابر با محدودیت همراه است. از سوی دیگر، حصول آگاهی از الگوهای مکانی- زمانی گردوغبار در آینده برای اتخاذ تدابیر مدیریت بحران حائز اهمیت است. با به‌کارگیری مدل‌های پیش‌بینی‌کننده استخراج‌شده در این مطالعه می‌توان در مکان‌های فاقد AOD، برآورد قابل قبولی از AOD به‌عمل آورد. همچنین، با واردسازی مقادیر پیش‌بینی‌شده پارامترهای اقلیمی، می‌توان الگوهای مکانی- زمانی AOD را در آینده نیز تخمین زد.

رخداد طوفان‌های گردوغبار عموماً تابعی از طیف وسیعی از شرایط محیطی شامل خصوصیات اتمسفری و همچنین پارامترهای سطح زمین نظیر پوشش گیاهی، رطوبت، و بافت خاک است. با این علم، صرف در نظر گرفتن وضعیت اتمسفر و تأثیر آن بر الگوهای مکانی- زمانی AOD ممکن است در برخی مواقع نتایج مطلوبی به‌دست ندهد. بنابراین، توصیه می‌شود در مطالعات آتی در این زمینه، علاوه بر پارامترهای اقلیمی که بیشتر نمایه‌هایی از وضعیت اتمسفرند، از پارامترهای سطح زمین نیز در مدل‌سازی استفاده شود. با انجام‌دادن این کار انتظار می‌رود میزان صحت مدل‌های خطی برای پیش‌بینی AOD افزایش یابد.

## منابع

- Alsultanny, Y. (2020). Machine Learning by Data Mining REPTree and M5P for Predicating Novel Information for PM10. *Cloud Computing and Data Science*, 40-48.
- Andina, D. and Pham, D. T. (2007). *Computational intelligence For engineering and manufacturing*. Springer.
- Bellinger, C.; Jabbar, M. S. M.; Zaïane, O. and Osornio-Vargas, A. (2017). A systematic review of data mining and machine learning for air pollution epidemiology. *BMC Public Health*, 17(1): 1-19.
- Chu, D. A.; Kaufman, Y. J.; Ichoku, C.; Remer, L. A.; Tanré, D. and Holben, B. N. (2002). Validation of MODIS aerosol optical depth retrieval over land. *Geophysical Research Letters*, 29(12): MOD2-1.
- Darvishi Bolorani, A.; Papi, R.; Soleimani, M.; Karami, L.; Amiri, F. and Samany, N. N. (2021). Water bodies changes in Tigris and Euphrates basin has impacted dust storms phenomena. *Aeolian Research*, 50: 100698.
- Du, M.; Yonemura, S.; Den, H.; Shen, Z. and Shen, Y. (2009). Relationship between the climate change and dust storm occurrence in China. *J Arid Land Stud*, 19(1): 149-152.
- Frank, E.; Wang, Y.; Inglis, S.; Holmes, G. and Witten, I. H. (1998). Using model trees for classification. *Machine Learning*, 32(1): 63-76.
- Gholami, H.; Mohamadifar, A. and Collins, A. L. (2020). Spatial mapping of the provenance of storm dust: Application of data mining and ensemble modelling. *Atmospheric Research*, 233: 104716.
- Ghorbanzadeh, O.; Rostamzadeh, H.; Blaschke, T.; Gholaminia, K. and Aryal, J. (2018). A new GIS-based data mining technique using an adaptive neuro-fuzzy inference system (ANFIS) and k-fold cross-validation approach for land subsidence susceptibility mapping. *Natural Hazards*, 94(2): 497-517.
- Ginoux, P.; Garbuzov, D. and Hsu, N. C. (2010). Identification of anthropogenic and natural dust sources using Moderate Resolution Imaging Spectroradiometer (MODIS) Deep Blue level 2 data. *Journal of Geophysical Research Atmospheres*, 115(D5).
- Ginoux, P.; Prospero, J. M.; Gill, T. E.; Hsu, N. C. and Zhao, M. (2012). Global- scale attribution of anthropogenic and natural dust sources and their emission rates based on MODIS Deep Blue aerosol products. *Reviews of Geophysics*, 50(3).
- Goswami, S.; Chakraborty, S.; Ghosh, S.; Chakrabarti, A. and Chakraborty, B. (2018). A review on application of data mining techniques to combat natural disasters. *Ain Shams Engineering Journal*, 9(3): 365-378.
- Goudie, A. S. (2009). Dust storms: Recent developments. *Journal of Environmental Management*, 90(1): 89-94.
- Hsu, N. C.; Tsay, S.-C.; King, M. D. and Herman, J. R. (2004). Aerosol properties over bright-reflecting source regions. *IEEE Transactions on Geoscience and Remote Sensing*, 42(3): 557-569.
- Kaufman, Y. J.; Tanré, D.; Remer, L. A.; Vermote, E. F.; Chu, A. and Holben, B. N. (1997). Operational remote sensing of tropospheric aerosol over land from EOS moderate

- resolution imaging spectroradiometer. *Journal of Geophysical Research Atmospheres*, 102(D14): 17051-17067.
- King, M. D.; Kaufman, Y. J.; Menzel, W. P. and Tanre, D. (1992). Remote sensing of cloud, aerosol, and water vapor properties from the moderate resolution imaging spectrometer (MODIS). *IEEE Transactions on Geoscience and Remote Sensing*, 30(1): 2-27.
- Lee, S.; Lee, M.-J. and Jung, H.-S. (2017). Data mining approaches for landslide susceptibility mapping in Umyeonsan, Seoul, South Korea. *Applied Sciences*, 7(7): 683.
- Li, C.; Lau, A.-H.; Mao, J. and Chu, D. A. (2005). Retrieval, validation, and application of the 1-km aerosol optical depth from MODIS measurements over Hong Kong. *IEEE Transactions on Geoscience and Remote Sensing*, 43(11): 2650-2658.
- Nabavi, S. O.; Haimberger, L. and Samimi, C. (2016). Climatology of dust distribution over West Asia from homogenized remote sensing data. *Aeolian Research*, 21: 93-107.
- Najafi, M. S.; Khoshakhllagh, F.; Zamanzadeh, S. M.; Shirazi, M. H.; Samadi, M. and Hajikhani, S. (2014). Characteristics of TSP Loads during the Middle East Springtime Dust Storm (MESDS) in Western Iran. *Arabian Journal of Geosciences*, 7(12): 5367-5381. <https://doi.org/10.1007/s12517-013-1086-z>
- Namdari, S.; Karimi, N.; Sorooshian, A.; Mohammadi, G. and Sehatkashani, S. (2018). Impacts of climate and synoptic fluctuations on dust storm activity over the Middle East. *Atmospheric Environment*, 173: 265-276.
- Oprea, M.; Dragomir, E. G.; Popescu, M. and Mihalache, S. F. (2016). Particulate matter air pollutants forecasting using inductive learning approach. *Rev. Chim.*, 67: 2075-2081.
- Pal, M. (2006). M5 model tree for land cover classification. *International Journal of Remote Sensing*, 27(4): 825-831.
- Prospero, J. M.; Ginoux, P.; Torres, O.; Nicholson, S. E. and Gill, T. E. (2002). Environmental characterization of global sources of atmospheric soil dust identified with the Nimbus 7 Total Ozone Mapping Spectrometer (TOMS) absorbing aerosol product. *Reviews of Geophysics*, 40(1): 1-2.
- Rahimikhoob, A.; Asadi, M. and Mashal, M. (2013). A comparison between conventional and M5 model tree methods for converting pan evaporation to reference evapotranspiration for semi-arid region. *Water Resources Management*, 27(14): 4815-4826.
- Rahmati, O.; Mohammadi, F.; Ghiasi, S. S.; Tiefenbacher, J.; Moghaddam, D. D.; Coulon, F.; ... and Bui, D. T. (2020). Identifying sources of dust aerosol using a new framework based on remote sensing and modelling. *Science of the Total Environment*, 737: 139508.
- Samadi, M.; Darvishi Bolorani, A.; Alavipanah, S.; Mohamadi, H. and Najafi, M. (2014). Global dust Detection Index (GDDI); a new remotely sensed methodology for dust storms detection. *Journal of Environmental Health Science and Engineering*, 12(1): 20. <https://doi.org/10.1186/2052-336X-12-20>
- Sayer, A. M.; Hsu, N. C.; Bettenhausen, C. and Jeong, M. (2013). Validation and uncertainty estimates for MODIS Collection 6 "Deep Blue" aerosol data. *Journal of Geophysical Research Atmospheres*, 118(14): 7864-7872.
- Schober, P.; Boer, C. and Schwarte, L. A. (2018). Correlation coefficients: appropriate use and interpretation. *Anesthesia & Analgesia*, 126(5): 1763-1768.

- Shaban, K. B.; Kadri, A. and Rezk, E. (2016). Urban air pollution monitoring system with forecasting models. *IEEE Sensors Journal*, 16(8): 2598-2606.
- Siwek, K. and Osowski, S. (2016). Data mining methods for prediction of air pollution. *International Journal of Applied Mathematics and Computer Science*, 26(2): 467-478.
- Srinivasan, D. B. and Mekala, P. (2014). Mining social networking data for classification using reptree. *International Journal of Advance Research in Computer Science and Management Studies*, 2(10).
- Tan, F.; San Lim, H.; Abdullah, K. and Holben, B. (2016). Estimation of aerosol optical depth at different wavelengths by multiple regression method. *Environmental Science and Pollution Research*, 23(3): 2735-2748.
- Trigo, R. M.; Gouveia, C. M. and Barriopedro, D. (2010). The intense 2007–2009 drought in the Fertile Crescent: Impacts and associated atmospheric circulation. *Agricultural and Forest Meteorology*, 150(9): 1245-1257.
- Willmott, C. J. and Matsuura, K. (2005). Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. *Climate Research*, 30(1), 79–82.
- Witten, I. H. and Frank, E. (2002). Data mining: practical machine learning tools and techniques with Java implementations. *Acm Sigmod Record*, 31(1): 76-77.
- Xu, H. and Deng, Y. (2017). Dependent evidence combination based on shearman coefficient and pearson coefficient. *IEEE Access*, 6: 11634-11640.
- Yu, Y.; Kalashnikova, O. V.; Garay, M. J.; Lee, H. and Notaro, M. (2018). Identification and characterization of dust source regions across North Africa and the Middle East using MISR satellite observations. *Geophysical Research Letters*, 45(13): 6690-6701.
- Zhao, C.; Dabu, X. and Li, Y. (2004). Relationship between climatic factors and dust storm frequency in Inner Mongolia of China. *Geophysical Research Letters*, 31(1).
- Zhou, Z.-H. (2003). Three perspectives of data mining. *Artificial Intelligence*, 143(1): 139-146. [https://doi.org/10.1016/S0004-3702\(02\)00357-0](https://doi.org/10.1016/S0004-3702(02)00357-0).