



# سنجش از دور & GIS ایران



سنجش از دور و GIS ایران سال پانزدهم، شماره چهارم، زمستان ۱۴۰۲  
Vol.15, No. 4, Winter 2023 Iranian Remote Sensing & GIS

۱۷-۳۰

مقاله پژوهشی

## بهبود دقت برآورد غلظت ازن در سطح زمین با استفاده از محصولات ماهواره‌ای و یادگیری ماشین

رسول آتشی<sup>۱</sup>، دلگانی<sup>۱</sup>، مینا مرادی زاده<sup>۲\*</sup>، بهنام تشیع<sup>۳</sup>

۱. دانشجوی کارشناسی ارشد، مهندسی نقشه‌برداری- گرایش سنجش از دور، دانشکده عمران

و حمل‌ونقل، دانشگاه اصفهان

۲. استادیار گروه مهندسی نقشه‌برداری، دانشکده عمران و حمل‌ونقل، دانشگاه اصفهان

تاریخ پذیرش مقاله: ۱۴۰۱/۰۶/۱۳

تاریخ دریافت مقاله: ۱۴۰۱/۰۴/۱۳

### چکیده

ازن نزدیک به سطح زمین یکی از آلاینده‌های بسیار خطرناک است که تأثیرات زیان‌بار درخور توجهی در سلامت ساکنان مناطق شهری دارد. هدف از این مطالعه شناسایی عوامل مؤثر در غلظت ازن و مدل‌سازی تغییرات آن، با استفاده از داده‌های ماهواره‌ای و روش‌های گوناگون یادگیری ماشین در شهر تهران است. بدین منظور داده‌های غلظت آلاینده‌ها، داده‌های هواشناسی و دمای سطح خاک، طی بازه زمانی بین سال‌های ۲۰۱۵ تا ۲۰۲۱، به کار رفت. پس از محاسبه همبستگی بین غلظت ازن و پارامترهای مستقل، طی پنج حالت متفاوت، با پارامترهای ورودی و روش یادگیری متفاوت و به کارگیری پالایش داده‌ها، غلظت ازن مدل‌سازی شد. در حالت اول و دوم، مدل‌سازی با استفاده از داده‌های غلظت آلاینده‌ها و داده‌های هواشناسی با روش رگرسیون خطی چندمتغیره انجام شد. تنها تفاوت این دو حالت، پالایش داده‌های ورودی به شیوه WTEST در روش دوم است. در حالت سوم، دمای سطح خاک به داده‌های ورودی افزوده شد و در حالت چهارم و پنجم، به ترتیب مدل‌سازی ازن با استفاده از شبکه عصبی چندلایه‌ای و شبکه عصبی بازگشتی انجام شد. مقایسه این حالت‌ها نشان داد که مدل‌سازی‌های مراحل اول تا پنجم، به ترتیب با ضریب تعیین تعدیل‌شده ۰/۵، ۰/۶۴، ۰/۷۴، ۰/۸ و توانایی بازایی غلظت ازن را داشته‌اند. همچنین مشخص شد در بین آلاینده‌های گوناگون، مونوکسید نیتروژن، دی‌اکسید نیتروژن، نیتراکس و از میان داده‌های هواشناسی دما، رطوبت و سرعت باد بیشترین تأثیر را در غلظت ازن دارند. افزودن دمای سطح خاک به داده‌های ورودی نیز افزایش پنج‌درصدی دقت را در برآورد غلظت ازن، به همراه داشت.

**کلیدواژه‌ها:** غلظت ازن، یادگیری ماشین، رگرسیون خطی چندمتغیره، شبکه عصبی بازگشتی، آلاینده جوئی.

\* نویسنده مکاتبه‌کننده: اصفهان، میدان آزادی، دانشگاه اصفهان، دانشکده فنی و مهندسی، گروه مهندسی نقشه‌برداری، تلفن: ۰۳۱۳۷۹۳۵۳۱۲

Email: m.moradzadeh@eng.ui.ac.ir

<https://orcid.org/0000-0003-3163-5853>

<https://dx.doi.org/10.48308/gisj.2022.102758>



Copyright: © 2023 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## ۱- مقدمه

عملکردهای تنفسی، در درازمدت باعث ازکارافتادن ریه‌ها می‌شود (Ito et al., 2005). از مضرات گاز ازن در سطح زمین می‌توان به تأثیر گسترده آن در اکوسیستم‌های زمینی، افزایش اثر گلخانه‌ای و افزایش دمای کره زمین و کاهش تولید محصولات کشاورزی اشاره کرد (Hollaway et al., 2012). ازن، از طریق روزنه، به برگ‌های گیاه وارد می‌شود و با اکسید کردن بافت گیاه، عملیات فتوسنتز را مختل می‌کند (Booker et al., 2009). آلاینده‌های هوا به‌طور کلی در چهار دسته قرار می‌گیرد:

- ۱) آلاینده‌های گازی (از جمله ازن، دی‌اکسید سولفور<sup>۲</sup>، دی‌اکسید نیتروژن<sup>۳</sup>، مونوکسید نیتروژن<sup>۴</sup>، دی‌اکسید کربن<sup>۵</sup>، مونوکسید کربن<sup>۶</sup> و ترکیبات آلی فرآر)؛
- ۲) آلاینده‌های آلی دیرپا (مانند دیوکسین<sup>۷</sup>)؛
- ۳) فلزات سنگین (مثلاً سرب و جیوه)؛
- ۴) ذرات معلق<sup>۸</sup> (Kampa & Castanas, 2008).

عبدالوهاب<sup>۹</sup> و همکاران (۲۰۰۵) با مدل رگرسیون خطی چندمتغیره<sup>۱۰</sup> و به‌کارگیری پارامترهای مربوط به آلاینده‌های جوّی و پارامترهای هواشناسی مقدار غلظت ازن را پیش‌بینی کردند و به این نتیجه رسیدند که گازهای دی‌اکسید نیتروژن، مونوکسید نیتروژن و دما بیشترین تأثیر را در افزایش غلظت ازن دارند. شیانگ<sup>۱۱</sup> و همکاران (۲۰۲۰)، با مقایسه روش‌های رگرسیون خطی و الگوریتم‌های یادگیری ماشین<sup>۱۲</sup>، غلظت ازن را با استفاده از داده‌های

وجود و پخش یک یا چند آلوده‌کننده اعم از جامد، مایع، گاز، تشعشع پرتوزا و غیرپرتوزا در هوای آزاد، به‌مقدار و مدتی که کیفیت هوا را تغییر دهد و آن را برای انسان و سایر موجودات زنده یا گیاهان و یا آثار تاریخی و بناها زیان‌آور کند، آلودگی هوا محسوب می‌شود (Webster, 2014). یکی از اصلی‌ترین نگرانی‌ها و مشکلات زیست‌محیطی، در دهه‌های اخیر، افزایش مداوم سطح گازهای آلاینده در نزدیکی سطح زمین است (Javanbakht Amiri & Khatami., 2012). براساس گزارش سازمان بهداشت جهانی در سال ۲۰۱۲، هر سال بیش از ۳/۷ میلیون نفر بر اثر بیماری‌های ناشی از آلودگی هوا جان خود را از دست می‌دهند (Asl et al., 2018). در طول قرن اخیر، غلظت گاز ازن<sup>۱</sup> در سطح زمین دوبرابر شده و این درحالی است که مقدار زیاد این آلاینده برای سلامت انسان بسیار خطرناک‌تر از دیگر آلاینده‌هاست و از عوامل اصلی بیماری‌های ریوی به‌شمار می‌رود (Ito et al., 2005).

تروپوسفر پایین‌ترین لایه اتمسفر زمین است. این لایه تا ارتفاع ده کیلومتری سطح زمین ادامه می‌یابد. لایه استراتوسفر از لایه تروپوسفر آغاز می‌شود و تا پنجاه کیلومتری سطح زمین ادامه دارد. لایه ازن در استراتوسفر وجود دارد. مولکول‌های ازن، در این لایه، انرژی فرابنفش خورشید را جذب و آن را به گرما تبدیل می‌کنند. ازن موجود در لایه استراتوسفر از انسان در برابر اشعه فرابنفش خورشید محافظت می‌کند. ازن تروپوسفر که نام دیگر آن ازن سطح زمین است، دو منبع دارد:

- ۱) ریزش ازن استراتوسفر به جوّ و تبدیل آن به ازن سطح زمین؛
  - ۲) واکنش شیمیایی بین برخی آلاینده‌ها و تبدیل آن به ازن سطح زمین.
- غلظت بالای این آلاینده تأثیرات بسیار نامطلوبی در سلامت انسان دارد (Susaya et al., 2013). ازن با تحریک ریه و تأثیر در بافت‌های آن، غشای مخاطی و

1. Ozone (O3)
2. Sulfur Dioxide (SO2)
3. Nitrogen Dioxide (NO2)
4. Nitric Monoxide (NO)
5. Carbon Dioxide (CO2)
6. Carbon Monoxide (CO)
7. Dioxins
8. Particulate Matter (PM)
9. Abdul-Wahab
10. Multivariate Linear Regression (MLR)
11. Xiang
12. Machine Learning

کنترلی و مدیریتی، ضرورت می‌یابد. طی دهه‌های گذشته، محققان روش‌های گوناگونی برای پیش‌بینی غلظت ازن مطرح کرده‌اند که می‌توان آنها را در دو دسته روش‌های تجربی و فیزیکی قرار داد. مدل‌های فیزیکی، به‌رغم دقت بالا، به‌دلیل حجم محاسبات و نیاز به پارامترهای اتمسفری و محیطی، پیچیدگی‌های زیادی دارند. این درحالی است که محققان بسیاری همواره روش‌های تجربی را به‌دلیل سادگی، استفاده کرده‌اند. درعین حال این روش‌ها قطعی نیستند و به‌کارگیری آنها، در شرایط مکانی و زمانی متفاوت، نیازمند فرایند یادگیری و کالیبراسیون دیگر باره است. روش‌های تجربی مبتنی بر رگرسیون خطی چندمتغیره، در مطالعات متعددی، به‌منظور پیش‌بینی غلظت ازن به‌کار رفته و توسعه یافته‌اند (Huang et al., 2019). ازسوی دیگر، به‌دلیل توانایی در انتخاب و استفاده از متغیرهای مستقل متعدد به‌منظور پیش‌بینی مطلوب متغیر وابسته، روش‌های یادگیری ماشین در سال‌های اخیر بسیار کاربرد داشته است. ازاین‌میان، مدل‌های مبتنی بر شبکه‌های عصبی<sup>۴</sup> همواره برای مدل‌سازی و ارائه مدل پیش‌بینی‌کننده‌ای قوی، از جایگاه ویژه‌ای برخوردار بوده‌اند (Biancofiore et al., 2015). در برخی مطالعات، بین کارایی روش‌های مبتنی بر رگرسیون خطی چندمتغیره و روش‌های مبتنی بر شبکه عصبی مقایسه انجام شده که نتایج آن به برتری عملکرد شبکه‌های عصبی اشاره داشته است (Spellman, 1999). توانایی مدل‌سازی رابطه‌ی غیرخطی بین ازن و پارامترهای وابسته در شبکه‌های عصبی باعث افزایش دقت مدل‌سازی تغییرات غلظت ازن و دلیل اصلی این برتری است. شایان ذکر است که در میان انواع گوناگون شبکه‌های عصبی، شبکه عصبی با استفاده از

ایستگاه‌های مانیتورینگ ازن و پارامترهای آلودگی هوا مدل‌سازی کردند و به این نتیجه رسیدند که الگوریتم‌های رگرسیون خطی جزء الگوریتم‌های ساده، قوی و سریع‌اند؛ درحالی که الگوریتم‌های یادگیری ماشین پیچیده‌تر و کندترند اما می‌توانند رابطه غیرخطی بین ازن و پارامترهای وابسته را مدل کنند که این نکته پتانسیل بهبود دقت مدل را افزایش می‌دهد (Ren et al., 2020). پاک<sup>۱</sup> (۲۰۱۸) در مطالعات خود به این نتیجه رسید که شبکه عصبی<sup>۲</sup> با استفاده از پارامترهای هواشناسی، به‌منزله داده ورودی، عملکرد پیش‌بینی‌کننده‌های عصبی را بهبود چشمگیر می‌بخشد (Pak et al., 2018).

دی‌اکسید نیتروژن، مونوکسید نیتروژن، دی‌اکسید کربن و مونوکسید کربن از گازهای اولیه در فرایند تولید ازن به‌شمار می‌روند و بدون این گازها، فرایند فتوشیمیایی انجام نمی‌شود (Ahmad et al., 2014). با برخورد اشعه خورشید به این گازها، پیوند مولکولی بین اتم‌های آنها شکسته می‌شود که به آزاد شدن اتم‌های اکسیژن می‌انجامد (Sullivan et al., 2017). این اتم‌ها با مولکول‌های اکسیژن<sup>۳</sup> واکنش می‌دهند و گاز ازن را به‌وجود می‌آورند (Reeves et al., 2002). پارامترهای هواشناسی رطوبت نسبی هوا و باد از دیگر عوامل مهم تأثیرگذار در غلظت ازن هستند. با افزایش رطوبت نسبی هوا، درجه حرارت و در نتیجه، مقدار ازن کاهش می‌یابد و باد نیز، به‌منزله عاملی مهم در پراکندگی آلودگی‌های جوئی، رابطه‌ای مستقیم با غلظت آلاینده ازن دارد (Sullivan et al., 2017). ازسوی دیگر، تغییرات مقدار غلظت ازن با تغییرات زمان و مکان، نشان می‌دهد که میزان فعالیت‌های انسانی در زمان‌های متفاوت، طول و عرض جغرافیایی و ارتفاع نیز نقش مهمی در میزان غلظت این پارامتر دارند (Vingarzan, 2004; Li & Rappenglück, 2014)؛ بنابراین، به‌دلیل تغییرات چرخه‌ای نامنظم ازن که ناشی از تأثیرپذیری از عوامل یادشده در بالاست، مدل‌سازی تغییرات ازن در نقاط و فصل‌های متفاوت به‌منظور بیان راهکارهای

1. Pak
2. Recurrent Neural Networks (RNN)
3. Molecular Oxygen (O2)
4. Perceptron

ارتفاع ۹۰۰ تا ۱۸۰۰ متری از سطح دریا امتداد یافته است. این شهر دارای اقلیم نیمه‌خشک است و در بیشتر سال‌ها، فصل زمستان نیمی از کل بارش‌های سالیانه تهران را تأمین می‌کند و تابستان نیز کم‌باران‌ترین فصل است (Sotoudehian & Arhami, 2014)؛ بنابراین پیش‌بینی غلظت ازن، با توجه به روند افزایشی آن در شهر تهران می‌تواند نقش مهمی در مدیریت آن ایفا کند (Ahmadi & Mahmoudi, 2013).



شکل ۱. محدوده شهر تهران به همراه ایستگاه‌های سنجش غلظت آلاینده‌ها و ایستگاه‌های سینوپتیک

## ۲-۲- داده‌ها

داده‌های مورد استفاده در این مطالعه شامل دما (برحسب درجه سلسیوس)، رطوبت نسبی (برحسب درصد) و سرعت باد (برحسب متر بر ثانیه) برگرفته از سازمان هواشناسی کشور، غلظت مونوکسید نیتروژن<sup>۲</sup> (برحسب ppb)، غلظت دی‌اکسید نیتروژن (برحسب ppb) و غلظت نیتراکس<sup>۳</sup> (برحسب ppb) برگرفته از شرکت کنترل کیفیت هوای تهران و همچنین محصول ماهواره‌ای دمای سطح خاک<sup>۴</sup> (برحسب کلین) است. در مدل‌سازی تغییرات غلظت ازن، استفاده از محصولات ماهواره‌ای به‌منزله یکی از اصلی‌ترین منابع

ترکیب پارامترهای ورودی، پتانسیل عملکرد پیش‌بینی‌کننده‌های عصبی را بهبود درخور توجهی بخشیده است (Pak et al., 2018). اگرچه دقت مدل‌سازی پارامترهای گوناگون علاوه بر الگوریتم مورد استفاده، تا حد بسیاری از رفتار پارامتر مورد بازیابی، مقیاس، وضوح و اندازه جامعه مورد مطالعه نشئت می‌گیرد، نقش کلیدی متغیرهای ورودی در مدل‌سازی تغییرات پارامتر مورد بازیابی، انکارناپذیر است. مطالعات نشان داده‌اند که گازهای دی‌اکسید نیتروژن، مونوکسید نیتروژن و دما بیشترین تأثیر را در تغییرات غلظت ازن داشته‌اند (Abdul-Wahab et al., 2005). هدف از این پژوهش مدل‌سازی غلظت ازن، با استفاده از غلظت آلاینده‌های جوئی و داده‌های هواشناسی است. این داده‌ها، به‌ترتیب، از ایستگاه‌های سنجش آلاینده‌های جوئی و ایستگاه‌های سینوپتیک شهر تهران، طی بازه زمانی سال‌های ۲۰۱۵ تا ۲۰۲۱ تهیه شده‌اند. از آنجاکه دمای سطح زمین همبستگی بسیاری با غلظت ازن سطحی دارد و محصول ماهواره‌ای دمای سطح خاک<sup>۱</sup> نیز با توان تفکیک مکانی ۱۰۰۰ متر، به‌صورت روزانه در دسترس قرار دارد، در این مطالعه به‌منظور بهبود دقت مدل‌سازی، پارامتر مورد اشاره به داده‌های ورودی افزوده شده است (Sullivan et al., 2017).

## ۲- مواد و روش‌ها

### ۲-۱- منطقه مطالعاتی

منطقه مورد مطالعه، در این تحقیق، شهر تهران با محدوده طول و عرض جغرافیایی نشان‌داده‌شده در شکل ۱ است. تهران یکی از آلوده‌ترین شهرهای جهان است که تراکم جمعیتی بالا، طوفان‌های گردوغبار، توپوگرافی خاص و احاطه شهر با کوه‌های اطراف از عوامل اصلی آلودگی‌اند. این شهر، با ۸،۶۹۳،۷۰۶ نفر جمعیت، بیست و چهارمین شهر پرجمعیت جهان و دومین کلان‌شهر پرجمعیت خاورمیانه است. از دید ناهمواری‌های طبیعی، شهر تهران به دو ناحیه دشتی و کوهپایه‌ای البرز تقسیم می‌شود و گستره کنونی آن از

1. LST MOD11A1
2. Parts-Per-Billion (ppb, 10-9)
3. Nitrox (NOx)
4. MODIS/ Terra Land Surface Temperature/Emissivity Daily L3 Global 1 km

بین ازن و دما برقرار می‌شود. بررسی سری زمانی سه آلاینده دی‌اکسید نیتروژن، مونوکسید نیتروژن و مونوکسید کربن که از گازهای اولیه در فرایند تولید ازن محسوب می‌شوند، حاکی از ارتباط غیرمستقیم تغییرات غلظت این آلاینده‌ها با آلاینده ازن بوده است (Ahmad et al., 2014). بنابراین در مرحله اول تحقیق، میزان همبستگی بین این متغیرها با مقدار ازن باید بررسی و ارزیابی شود. جدول ۱ توصیف آماری ۲۲۸۰۰ داده (داده‌های مربوط به بازه زمانی پنج‌ساله از ایستگاه‌های آلاینده) را نشان می‌دهد. میزان غلظت ازن از ۱ تا ۸۳ متغیر بوده و دارای میانگین ۲۱/۷ است. نسبت میانگین به دامنه تغییرات، در این آلاینده، بیشتر از دیگر آلاینده‌هاست که مقدار زیاد این آلاینده را می‌رساند. میانگین رطوبت کمتر از ۵۰ و دمای رو به بالا بیانگر شرایط آب‌وهوایی خشک و گرم است.

میزان ضریب همبستگی پیرسون بین پارامترهای مستقل و غلظت ازن که از طریق رابطه (۲) محاسبه می‌شود، در جدول ۲ نشان داده شده است (Özbay, 2012).

$$R = \frac{\sum(X-\bar{X})(Y-\bar{Y})}{\sqrt{\sum(X-\bar{X})^2 \sum(Y-\bar{Y})^2}} \quad (2)$$

در این رابطه،  $Y, X, \bar{Y}$  و  $\bar{X}$  به ترتیب بردار متغیر مستقل، بردار متغیر وابسته، بردار میانگین متغیر مستقل و بردار میانگین متغیر وابسته‌اند.

تأمین متغیرهای ورودی، موجب افزایش دقت پیش‌بینی پارامتر ازن می‌شود (Abdullah et al., 2019). داده‌های ایستگاه‌های سینوپتیک هر سه ساعت یک‌بار برداشت می‌شود و از وبگاه <https://data.irimo.ir/> می‌توان آن را دانلود کرد. این داده‌ها با استفاده از میانگین‌گیری ساده به داده‌های روزانه تبدیل شده و در مدل‌سازی به کار رفته‌اند. داده‌های ایستگاه‌های آلاینده نیز روزانه از وبگاه <https://airnow.tehran.ir/home/DataArchive.aspx> دانلود شده‌اند. داده‌های ایستگاه‌های آلاینده و سینوپتیک در بازه زمانی سال‌های ۲۰۱۵ تا ۲۰۲۰ (پنج سال)، با توان تفکیک روزانه، دریافت و استفاده شدند.

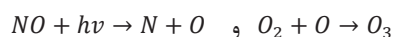
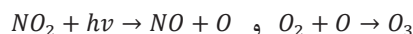
### ۳-۲- روش کار

#### ۳-۳-۱- توصیف آماری داده‌ها و محاسبه همبستگی

##### بین غلظت ازن با متغیرهای مستقل

همان‌طور که اشاره شد، گازهای مونوکسید نیتروژن و دی‌اکسید نیتروژن از گازهای اصلی در فرایند تولید ازن شمرده می‌شوند (Ahmad et al., 2014). رابطه (۱) فرایند تشکیل این گاز را بیان می‌کند.

رابطه (۱)



در این رابطه، مقدار  $hv$  برابر با انرژی اشعه خورشید است و به این ترتیب، از طریق اشعه خورشید، ارتباطی

جدول ۱. توصیف آماری پارامترهای وابسته و مستقل استفاده‌شده در مدل‌سازی

غلظت ازن	غلظت مونوکسید نیتروژن	غلظت دی‌اکسید نیتروژن	غلظت نیتراکس	دمای سطح خاک	دمای هوا	رطوبت نسبی	سرعت باد
21/71	70/00	50/45	120/04	14805/20	18/86	34/09	2/11
11/30	56/95	24/08	71/69	646/49	9/89	16/89	1/27
1	0	0	0	12611	12/-6	7/874	0
83	459	358	644	16353	6/37	100	10/5

جدول ۲. مقادیر ضریب همبستگی ازن و پارامترهای مستقل

سرعت باد	رطوبت نسبی	دمای هوا	دمای سطح خاک	غلظت نیتراکس	غلظت دی‌اکسید نیتروژن	غلظت مونوکسید نیتروژن
0/14	-0/51	0/63	0/65	-0/40	-0/22	-0/41
<0/001	<0/001	<0/001	<0/001	<0/001	<0/001	<0/001

خطای تصادفی رگرسیون است. با بازنویسی رابطه (۳) براساس پارامترهای ورودی حالت اول، رابطه (۴) حاصل خواهد شد که مدل سازی براساس آن انجام شده است.

رابطه (۴)

$$O_{ij} = b_0 + (b_1 NO_{2ij}) + (b_2 NO_{ij}) + (b_3 NO_{xij}) + (b_4 WS_{ij}) + (b_5 Temp_{ij}) + (b_6 RH_{ij}) + \varepsilon_{ij}$$

در این معادله  $NO_{2ij}$ ,  $NO_{ij}$ ,  $NO_{xij}$ ,  $Temp_{ij}$ ,  $RH_{ij}$ ,  $WS_{ij}$ ,  $\varepsilon_{ij}$  و  $O_{ij}$  به ترتیب، غلظت آلاینده های نیتراکس، دی اکسید نیتروژن، مونوکسید نیتروژن، دما، رطوبت نسبی، سرعت باد، خطای مدل سازی و غلظت ازن برداشت شده از ایستگاه  $i$  در روز  $j$  ام است.

در حالت دوم، باز هم از غلظت آلاینده ها و داده های هواشناسی به منزله پارامترهای ورودی روش رگرسیون خطی چندمتغیره استفاده شده؛ با این تفاوت که داده ها به روش WTEST نیز پالایش شده است. این آماره، با استفاده از بردار خطای باقی مانده و ماتریس واریانس-کوواریانس بردار خطای باقی مانده، مشاهدات اشتباه را شناسایی و حذف می کند و خطای مدل سازی را کاهش می دهد. در این روش، ابتدا بردار خطای باقی مانده رگرسیون و ماتریس واریانس-کوواریانس بردار خطا محاسبه می شود؛ سپس، با تقسیم عناصر خطای باقی مانده رگرسیون بر عناصر قطری ماتریس واریانس-کوواریانس بردار خطا، داده ای که بیشترین خطا را در مدل سازی ایجاد کرده است شناسایی و حذف می شود.

مطابق رابطه (۵)، تنها تفاوت حالت سوم مدل سازی ازن با حالت پیشین افزودن متغیر محصول ماهواره ای دمای سطح خاک، در کنار سایر پارامترهای ورودی، به مدل است.

رابطه (۵)

$$O_{ij} = b_0 + (b_1 NO_{2ij}) + (b_2 NO_{ij}) + (b_3 NO_{xij}) + (b_4 WS_{ij}) + (b_5 Temp_{ij}) + (b_6 RH_{ij}) + (b_7 LST_{ij}) + \varepsilon_{ij}$$

در این معادله،  $LST_{ij}$  دمای سطح زمین در پیکسل مربوط به ایستگاه  $i$  در روز  $j$  ام است.

نتایج همبستگی بین ازن و متغیرهای مستقل نشان می دهد، اگرچه گازهای مونوکسید نیتروژن و دی اکسید نیتروژن گازهای تولید ازن هستند، با غلظت آن همبستگی منفی دارند. تمایل بیشتر اتم اکسیژن برای واکنش با گازهای نیتروژن و مونوکسید نیتروژن در قیاس با اکسیژن ( $O_2$ ) دلیل این نکته است. البته با افزایش دما و متعاقب آن، افزایش احتمال شکستن پیوند مولکولی بین این گازها و آزاد شدن اکسیژن، فرایند تشکیل ازن تسریع می شود. از طرفی، با افزایش رطوبت نسبی، اثر دما و نور خورشید کاهش می یابد و فرایند تشکیل ازن کندتر می شود؛ به همین دلیل، ازن با رطوبت همبستگی منفی و با دما بیشترین همبستگی را دارد. از این رو در مناطق با دمای بیشتر، احتمال بالاتر بودن غلظت ازن بسیار بیشتر خواهد بود. مقادیر همبستگی و معناداری آماری ( $P < 0.01$ ) نشان می دهد همه متغیرهای مستقل به کاررفته در مدل سازی ها به افزایش دقت مدل سازی ازن منجر می شوند و عملکرد مدل ها را بهبود خواهند بخشید.

### ۲-۳-۲- مدل سازی

در این تحقیق، طی پنج حالت متفاوت و با استفاده از داده های در دست، مدل سازی غلظت ازن انجام شده است.

در حالت اول که غلظت آلاینده ها و داده های هواشناسی به منزله پارامترهای ورودی اند، به منظور مدل سازی، از روش رگرسیون خطی چندمتغیره استفاده شده است. یکی از تکنیک های رایج، در مدل سازی آلاینده ها، رگرسیون خطی چندمتغیره است که هدف آن کمینه کردن مجموع مربعات اختلاف ازن محاسبه شده و ازن واقعی است. فرمول کلی رگرسیون چندمتغیره مطابق رابطه (۳) است (Gvozdić et al., 2011).

$$y = b_0 + \sum_{i=1}^p b_i * x_i + \varepsilon \quad \text{رابطه (۳)}$$

در این معادله،  $y$  ازن برآورد شده،  $x_i$  متغیر مستقل،  $b_0$  عرض از مبدأ،  $b_i$  ضریب متغیرهای مستقل و  $\varepsilon$

مطابق روابط (۶) تا (۱۱)، ضریب تعیین مدل<sup>۱</sup>، جذر میانگین مربعات خطاها<sup>۲</sup>، میانگین مطلق خطاها<sup>۳</sup>، میانگین مربعات خطاها<sup>۴</sup>، جذر میانگین مربعات خطاهای نرمال شده<sup>۵</sup> و ضریب تعیین تعدیل شده مدل<sup>۶</sup> ملاک ارزیابی دقت و عملکرد مدل سازی های گوناگون قرار گرفتند (UI-Saufie et al., 2013).

$$R^2 = 1 - \frac{\sum(O_i - P_i)^2}{\sum(O_i - O_m)^2} \quad \text{رابطه (۶)}$$

در این معادله،  $R^2$  ضریب تعیین،  $O_i$  مقدار غلظت ازن اندازه گیری شده،  $n$  تعداد دفعات اندازه گیری ازن،  $P_i$  مقدار غلظت ازن مدل سازی،  $S_{obs}$  انحراف معیار استاندارد غلظت ازن،  $S_{pred}$  انحراف معیار غلظت ازن تخمین زده شده،  $\bar{P}$  میانگین غلظت ازن تخمین زده شده و  $\bar{O}$  میانگین غلظت ازن اندازه گیری شده است.

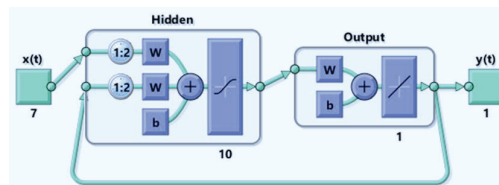
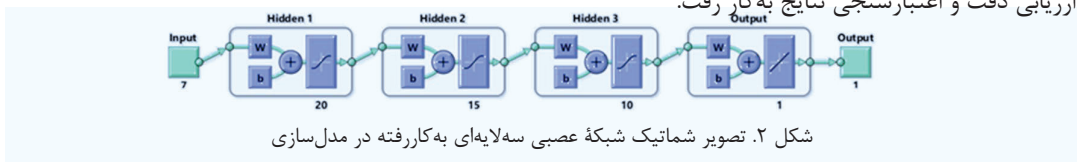
$$R^2 Ad = 1 - \frac{(1-R^2)(n-1)}{n-k-1} \quad \text{رابطه (۷)}$$

$R^2 Ad$  ضریب تعیین تعدیل شده،  $n$  تعداد مقادیر ازن اندازه گیری شده و  $k$  تعداد متغیرهای مستقل در مدل سازی است.

در حالت چهارم از غلظت آلاینده‌ها، داده‌های هواشناسی و محصول ماهواره‌ای دمای سطح خاک با به کارگیری روش شبکه عصبی چندلایه‌ای به همراه پالایش داده‌ها، به منظور مدل سازی ازن استفاده شد. در این حالت، شبکه عصبی سه لایه‌ای با تعداد نورون‌های ۱۰، ۱۵، ۲۰ در لایه‌های مخفی و توابع فعال سازی تانژانتی به کار رفت و آموزش شبکه براساس الگوریتم Levenberg Marqua انجام شد. دیاگرام شبکه استفاده شده مطابق شکل ۲ است. شبکه عصبی، به همراه پالایش داده، برای آموزش در حالت پنجم استفاده شد و داده‌های ورودی شامل غلظت آلاینده‌ها، داده‌های هواشناسی و محصول ماهواره‌ای دمای سطح خاک بود. آموزش شبکه عصبی تک لایه با ده نورون و تابع فعال سازی تانژانتی و میزان دمای ۲، براساس الگوریتم Levenberg Marqua انجام شد. دیاگرام شبکه به کاررفته در شکل ۳ نشان داده شده است.

### ۳-۳-۲- اجرای مدل و ارزیابی نتایج

به منظور مدل سازی برآورد غلظت ازن در حالت‌های متفاوت، ۷۰٪ داده‌ها استفاده شد و ۳۰٪ دیگر در ارزیابی دقت و اعتبارسنجی نتایج به کار رفت.



1. R-Squared (R2)
2. Root Mean Square Error (RMSE)
3. Mean Absolute Error (MAE)
4. Mean Square Error (MSE)
5. Normalized Root Mean Square Error (NRMSE)
6. Adjusted R-Squared (A\_R2)

$$MAE = \frac{1}{n} \sum |O_i - P_i| \quad \text{رابطه (۸)}$$

$$MSE = \frac{1}{n} \sum (O_i - P_i)^2 \quad \text{رابطه (۹)}$$

$$RMSE = \sqrt{\frac{1}{n} \sum (O_i - P_i)^2} \quad \text{رابطه (۱۰)}$$

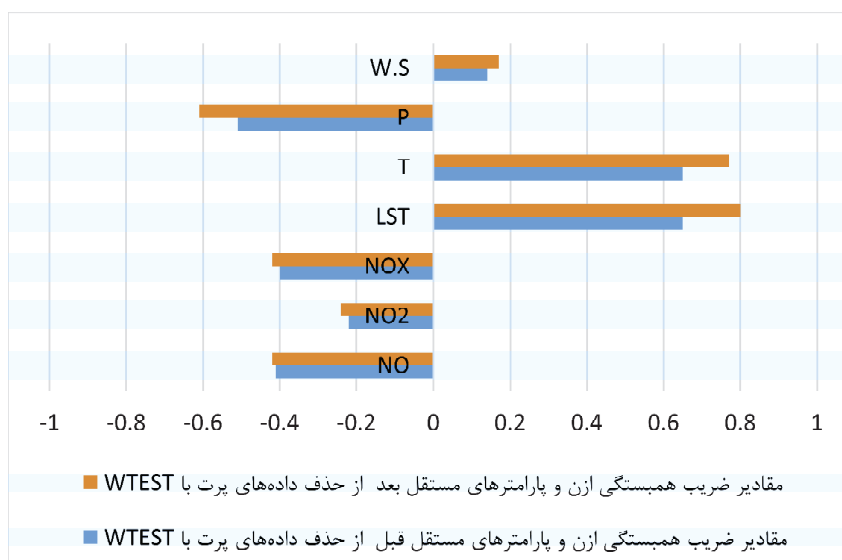
در معادلات بالا،  $MAE$  میانگین خطای مطلق،  $MSE$  میانگین مربعات خطا و  $RMSE$  جذر میانگین مربعات خطاست.

جذر میانگین مربعات خطای نرمال شده نیز طبق رابطه (۱۱) به دست می آید (Ul-Saufie et al., 2013).

$$NRMSE = \frac{RMSE}{O_{max} - O_{min}} \quad \text{رابطه (۱۱)}$$

### ۳- نتایج و بحث

در جدول ۳، مقادیر ضریب همبستگی بین ازن و پارامترهای مستقل، پس از حذف داده‌های پرت با  $WTEST$  آمده است. مقایسه مقادیر بیان شده در این جدول با جدول ۲ نشان می‌دهد حذف داده‌های پرت سبب افزایش شایان توجه همبستگی بین ازن و پارامترهای مستقل شده است. در شکل ۴، نتیجه اعمال پالایش و حذف داده‌های پرت و در جدول ۴، نتایج آماری عملکرد مدل‌سازی در حالت‌های متفاوت آمده است.



شکل ۴. مقایسه مقادیر ضریب همبستگی ازن و پارامترهای مستقل قبل و پس از حذف داده‌های پرت با تست  $WTEST$

جدول ۳. مقادیر ضریب همبستگی ازن و پارامترهای مستقل پس از حذف داده‌های پرت با تست  $WTEST$

	غلظت دی‌اکسید نیتروژن	غلظت مونوکسید نیتروژن	غلظت نیتراکس	دمای سطح خاک	دمای هوا	رطوبت نسبی	سرعت باد
ضریب همبستگی با ازن	-0/42	-0/24	-0/42	0/80	0/77	-0/61	0/17
مقادیر سطح معنی‌داری	<0/001	<0/001	<0/001	<0/001	<0/001	<0/001	<0/001

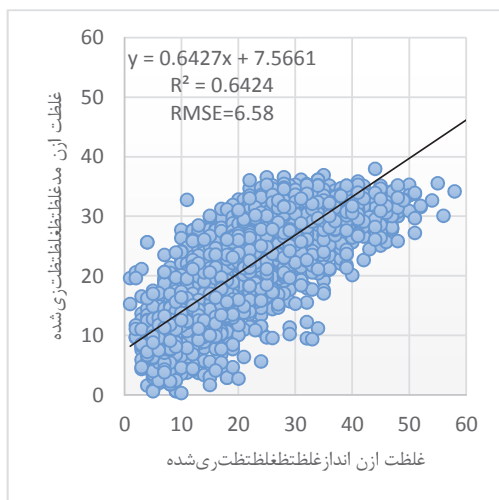


بهبود دقت برآورد غلظت ازن در سطح زمین ...

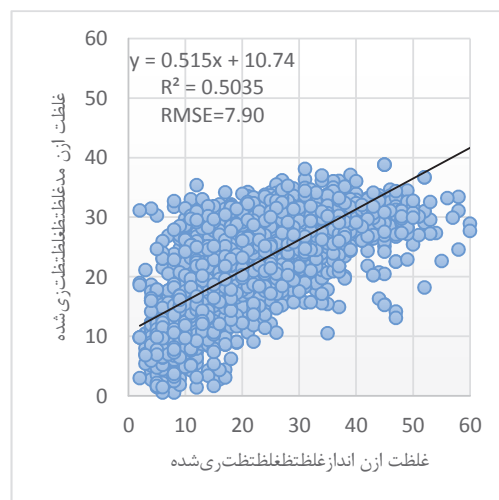
جدول ۴. نتایج آماری مدل‌سازی متغیر وابسته

پارامترهای ورودی	روش مدل‌سازی	پالایش داده	R2	Adjusted R2	MAE	MSE	RMSE	NRMSE
غلظت مونوکسید نیتروژن، غلظت دی‌اکسید نیتروژن، غلظت نیتراکس، دمای هوا، رطوبت نسبی، سرعت باد	برازش چندمتغیره	×	50/0	50/0	89/5	38/62	90/7	10/0
غلظت مونوکسید نیتروژن، غلظت دی‌اکسید نیتروژن، غلظت نیتراکس، دمای هوا، رطوبت نسبی، سرعت باد	برازش چندمتغیره	✓	64/0	64/0	11/5	24/43	58/6	08/0
غلظت مونوکسید نیتروژن، غلظت دی‌اکسید نیتروژن، غلظت نیتراکس، دمای هوا، رطوبت نسبی، سرعت باد، دمای سطح خاک	برازش چندمتغیره	✓	69/0	69/0	81/4	38/37	11/6	07/0
غلظت مونوکسید نیتروژن، غلظت دی‌اکسید نیتروژن، غلظت نیتراکس، دمای هوا، رطوبت نسبی، سرعت باد، دمای سطح خاک	شبکه عصبی پرسپترون چندلایه	✓	74/0	74/0	20/4	67/30	54/5	07/0
غلظت مونوکسید نیتروژن، غلظت دی‌اکسید نیتروژن، غلظت نیتراکس، دمای هوا، رطوبت نسبی، سرعت باد، دمای سطح خاک	شبکه عصبی	✓	80/0	80/0	98/3	78/25	21/5	06/0

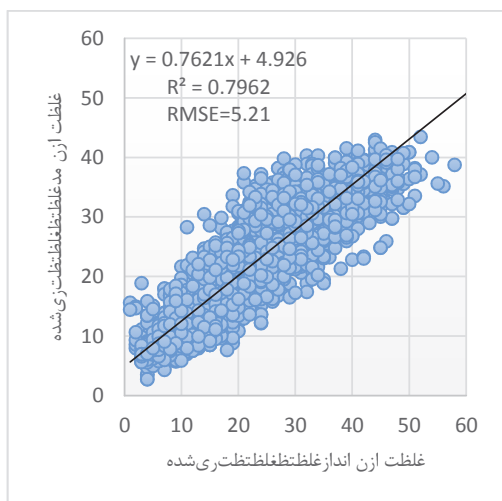
در شکل‌های ۵ تا ۹ نیز، نتایج آماری عملکرد مدل‌سازی در حالت‌های پنجگانه، در مقایسه با داده‌های غلظت ازن اندازه‌گیری‌شده، نشان داده شده است.



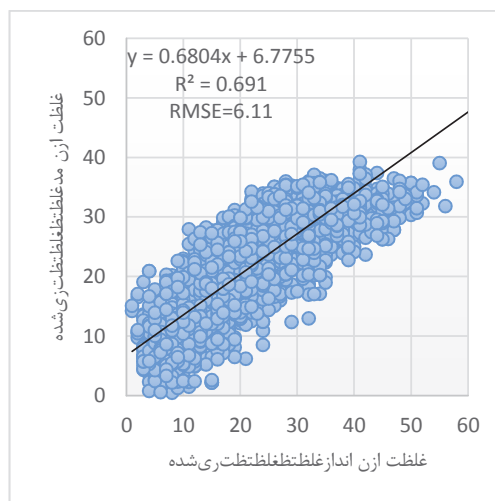
شکل ۶. مدل‌سازی با استفاده از غلظت آلاینده‌ها و داده‌های هواشناسی با روش رگرسیون خطی چندمتغیره و با پالایش داده‌ها به روش WTEST



شکل ۵. مدل‌سازی با استفاده از غلظت آلاینده‌ها و داده‌های هواشناسی با روش رگرسیون خطی چندمتغیره

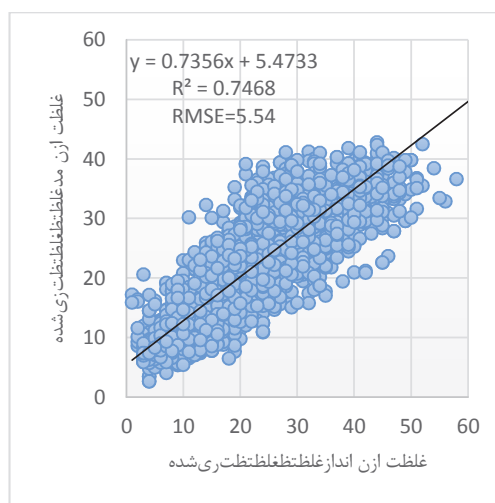


شکل ۹. مدل سازی با استفاده از غلظت آلاینده‌ها، داده‌های هواشناسی و محصول ماهواره‌ای دمای سطح خاک با روش شبکه عصبی و با پالایش داده‌ها به روش WTEST



شکل ۷. مدل سازی با استفاده از غلظت آلاینده‌ها، داده‌های هواشناسی و محصول ماهواره‌ای دمای سطح خاک با روش رگرسیون خطی چندمتغیره و با پالایش داده‌ها به روش WTEST

کار دقت مدل سازی را در مقایسه با حالت قبل، حدود ۱۸٪ بهبود بخشید ( $RMSE=6.5$ ,  $R^2=0.64$ ). با توجه به میزان همبستگی بالا و نقش تأثیرگذار دمای سطح خاک در برآورد میزان غلظت ازن، ورود محصول ماهواره‌ای دمای سطح خاک در کنار سایر پارامترها به مدل، در حالت سوم، توانست دقت مدل سازی را تا ۶٪ دیگر نیز بهبود دهد ( $RMSE=6.1$ ,  $R^2=0.69$ ). نتایج مدل سازی، در حالت چهارم، حاکی از آن است که شبکه عصبی پرسپترون چندلایه، در مقایسه با روش برازش چندمتغیره، توانایی بیشتری در برآورد غلظت ازن دارد و استفاده از آن می‌تواند خطای مدل سازی را به  $RMSE=5.5$  و ضریب تعیین بین مقادیر ازن برآورد شده و اندازه‌گیری شده را به  $R^2=0.7$  برساند. دلیل این موضوع توانایی شبکه عصبی پرسپترون چندلایه در کاهش داده‌های نویزدار در مدل سازی‌هاست (Hastie et al., 2009). شبکه‌های عصبی، به علت داشتن حافظه داخلی، می‌توانند ورودی قبلی خود را به خاطر بسپارند و از این حافظه برای پردازش دنباله‌ای از ورودی‌ها بهره ببرند؛ از این رو این شبکه‌ها توانایی مورد قبولی در مدل سازی داده‌های



شکل ۸. مدل سازی با استفاده از غلظت آلاینده‌ها، داده‌های هواشناسی و محصول ماهواره‌ای دمای سطح خاک با روش شبکه عصبی چندلایه‌ای و با پالایش داده‌ها به روش WTEST

در حالت اول، داده‌های ورودی برای مدل سازی غلظت ازن شامل برداشت‌های اغلب اشتباه و نویزدار می‌شود که به خطای بسیاری در مدل سازی منجر شده است ( $RMSE=7.9$ ,  $R^2=0.50$ )؛ بنابراین در حالت دوم، به منظور بهبود دقت مدل سازی، داده‌های اشتباه و نویزدار با استفاده از آماره WTEST حذف شد و این

روش‌های یادگیری ماشین مناسب، معرفی پارامترهای ورودی مرتبط و کاهش داده‌های نویزی، دقت و عملکرد مدل‌سازی بهبود یافته است؛ به گونه‌ای که این پژوهش، با به‌کارگیری ۲۲۸۰۰ داده در کل فرایند مدل‌سازی طی بازه زمانی سال‌های ۲۰۱۵ تا ۲۰۲۱، با استفاده از روش‌های مدل‌سازی بهتر و آماره WTEST، موفق به ارتقای مقادیر ضریب تعیین تعدیل‌شده از ۵۰ به ۸۰ شده است.

توزیع مکانی متوسط غلظت ازن اندازه‌گیری‌شده و مدل‌سازی‌شده پنج‌ساله (۲۰۱۵ تا ۲۰۲۰) به روش شبکه عصبی چندلایه‌ای در شکل ۱۰ نمایش داده شده است.

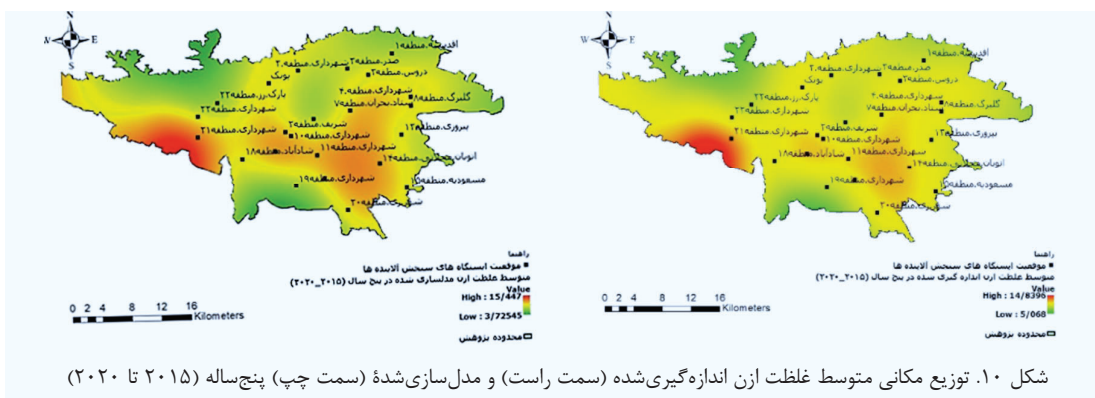
#### ۴- نتیجه‌گیری

با توجه به اهمیت غلظت ازن در جو و تأثیر آن در سلامت، رویکرد این تحقیق مدل‌سازی مناسب برای برآورد مداوم و دقیق این پارامتر جو بوده است. بدین منظور و با هدف معرفی قابلیت‌های محصولات ماهواره‌محور، از روش‌های شبکه عصبی همراه با داده‌های هواشناسی و آلاینده‌گی و محصول ماهواره‌ای دمای سطح خاک استفاده شد. در ابتداء، با استفاده از ضریب پیرسون، تأثیر هریک از متغیرهای مستقل و اهمیت آن در مدل‌سازی غلظت ازن بررسی شد. نتایج این همبستگی نشان داده است که تمامی پارامترهای مورد استفاده، به‌غیر از سرعت باد و دمای هوا،

سری زمانی دارند و می‌توانند عملکرد مدل‌سازی را بهبود بخشند. بنابراین در مدل‌سازی غلظت ازن در حالت پنجم، شبکه عصبی با همان پارامترهای ورودی مرحله پیشین استفاده شد و توانست دقت مدل‌سازی را حدود ۴٪ ارتقا دهد (RMSE=5.2, R<sup>2</sup>=0.80).

اگرچه برداشت نکردن غلظت آلاینده‌ها در برخی روزها و برداشت‌های اشتباه از جمله محدودیت‌ها و عوامل ایجاد خطا در این پژوهش بوده است، نتایج آماری، شامل ضریب تعیین بالا و RMSE پایین بین مقادیر اندازه‌گیری‌شده و مدل‌سازی‌شده غلظت ازن، با نتایج سایر پژوهش‌های گذشته، بیانگر توانایی شایان توجه مدل‌سازی انجام‌شده در پیش‌بینی غلظت ازن است (Draxler, 2000; Sousa et al., 2007; Shan et al., 2009; Barrero et al., 2006; Moustris et al., 2012).

همان‌طور که بررسی ضریب همبستگی پیرسون بین ازن و متغیرهای مستقل نشان داد، بیشترین همبستگی مثبت به دمای سطح زمین و بیشترین همبستگی منفی به غلظت مونوکسید نیتروژن متعلق است. بنابراین یکی از مهم‌ترین دلایل دستیابی به این سطح از دقت، ورود پارامتر دمای سطح خاک به‌منزله محصول ماهواره‌محور، همراه با داده‌های هواشناسی، به مدل پیش‌بینی‌کننده غلظت ازن است. می‌توان به‌طور کلی نتیجه گرفت که اگرچه همه مدل‌های توسعه‌یافته توانایی مدل‌سازی غلظت ازن را دارند، با مطرح‌کردن



جغرافیایی<sup>۱</sup> برای بهبود مدل سازی غلظت ازن استفاده کرد. رگرسیون وزنی جغرافیایی از روش های رگرسیون محلی و فضایی است که در مدل سازی روابط متغیرهای فضایی به کار می رود.

#### ۵- منابع

Abdul-Wahab, S.A., Bakheit, C.S. & Al-Alawi, S.M., 2005, **Principal Component and Multiple Regression Analysis in Modelling of Ground-Level Ozone and Factors Affecting Its Concentrations**, Environmental Modelling & Software, 20(10), PP. 1263-1271.

<https://doi.org/10.1016/j.envsoft.2004.09.001>

Abdullah, S., Nasir, N.H.A., Ismail, M., Ahmed, A.N. & Jarkoni, M.N.K., 2019, **Development of Ozone Prediction Model in Urban Area**, International Journal of Innovative Technology and Exploring Engineering, 8(10), PP. 2263-2267.

<https://doi.org/10.35940/ijitee.J1127.0881019>

Ahamad, F., Latif, M.T., Tang, R., Juneng, L., Dominick, D. & Juahir, H., 2014, **Variation of Surface Ozone Exceedance around Klang Valley**, Malaysia. Atmospheric Research, 139, PP. 116-127.

<https://doi.org/10.1016/j.atmosres.2014.01.003>

Ahmadi, M.M. & Mahmoudi, P., 2013, **Analysis of Tehran Air Pollution Data in Recent Decade (2000-2009)**, IJHE, 6(1), PP. 33-44.

Asl, F.B., Leili, M., Vaziri, Y., Arian, S.S., Cristaldi, A., Conti, G.O. & Ferrante, M., 2018, **Health Impacts Quantification of Ambient Air Pollutants Using AirQ Model Approach in Hamadan, Iran**, Environmental Research, 161, PP. 114-121.

<https://doi.org/10.1016/j.envres.2017.10.050>

Barrero, M., Grimalt, J.O. & Cantón, L., 2006, **Prediction of Daily Ozone Concentration Maxima in the Urban Atmosphere**, Chemometrics and Intelligent Laboratory Systems, 80(1), PP. 67-76.

<https://doi.org/10.1016/j.chemolab.2005.07.003>

همبستگی منفی با غلظت ازن داشته اند. غلظت ازن بیشترین همبستگی منفی را با غلظت مونوکسید نیتروژن (۴۱٪) و بیشترین همبستگی مثبت را با دمای هوا (مثبت ۶۳٪) دارد.

نتایج مدل سازی زمانی غلظت ازن نشان داد کمترین ضریب تعیین برای روش برازش خطی چندمتغیره (۵۰٪) و بیشترین ضریب تعیین برای روش شبکه عصبی (۸۰٪) است. از سویی، پالایش داده های ورودی باعث افزایش چهارده درصدی ضریب تعیین شد. افزودن محصول دمای سطح خاک سنجنده مادیس با مشاهدات زمینی نیز به افزایش پنج درصدی ضریب تعیین انجامید. با مقایسه نقشه غلظت ازن مدل سازی شده با نقشه غلظت ازن اندازه گیری شده از طریق ایستگاه های سینوپتیک، می توان نتیجه گرفت که مدل سازی های انجام شده تغییرات زمانی و مکانی غلظت ازن در شهر تهران را به خوبی برآورد کرده اند. پارامترهای مرتبط، به ویژه دمای سطح خاک، به منزله ورودی در مدل های شبکه عصبی معرفی شده وارد و با استفاده از مقادیر اندازه گیری شده ازن، مدل سازی انجام شد. در نهایت با استفاده از اندازه گیری های زمینی ازن، مدل ها اعتبارسنجی شدند و نتایج نشان داد که مدل سازی ها، در برآورد میزان ازن، قابلیت بالایی دارند. با توجه به احتمال تأثیر سایر پارامترهای سینوپتیک و آلودگی در غلظت ازن، مانند فشار هوا و غلظت مونوکسید کربن که به تغییرات گازهای جو منجر می شود، برای بررسی تأثیرات دیگر پارامترها در غلظت این پارامتر، مطالعات بیشتری ضرورت می یابد.

برای بهبود دقت مدل سازی می توان از افزودن محصولات ماهواره ای به داده های ورودی، مانند محصولات ماهواره ای وارونگی هوا و طیف UV-B خورشید که تغییرات ازن باعث تغییرات پیکسلی آن می شود، استفاده کرد.

از آنجا که مکان در مقدار غلظت ازن تأثیرگذار است، می توان از مدل های مکانی مانند مدل رگرسیون وزنی

1. Geographically Weighted Regression

- Biancofiore, F., Verdecchia, M., Di Carlo, P., Tomassetti, B., Aruffo, E., Busilacchio, M., Bianco, S., Di Tommaso, S. & Colangeli, C., 2015, **Analysis of Surface Ozone Using a Recurrent Neural Network**, Science of the Total Environment, 514, PP. 379-387.  
<https://doi.org/10.1016/j.scitotenv.2015.01.106>
- Booker, F., Muntifering, R., McGrath, M., Burkey, K., Decoteau, D., Fiscus, E., Manning, W., Krupa, S., Chappelka, A. & Grantz, D., 2009, **The Ozone Component of Global Change: Potential Effects on Agricultural and Horticultural Plant Yield, Product Quality and Interactions with Invasive Species**, Journal of Integrative Plant Biology, 51(4), PP. 337-351.  
<https://doi.org/10.1111/j.1744-7909.2008.00805.x>
- Draxler, R.R., 2000, **Meteorological Factors of Ozone Predictability at Houston, Texas**, Journal of the Air & Waste Management Association, 50(2), PP. 259-271.  
<https://doi.org/10.1080/10473289.2000.10463999>
- Gvozdić, V., Kovač-Andrić, E. & Brana, J., 2011, **Influence of Meteorological Factors NO<sub>2</sub>, SO<sub>2</sub>, CO and PM<sub>10</sub> on the Concentration of O<sub>3</sub> in the Urban Atmosphere of Eastern Croatia**, Environmental Modeling & Assessment, 16(5), PP. 491-501.  
<https://doi.org/10.1007/s10666-011-9256-4>
- Hastie, T., Tibshirani, R., Friedman, J.H. & Friedman, J.H., 2009, **The Elements of Statistical Learning: Data Mining, Inference, and Prediction** (Vol. 2), Springer.
- Hollaway, M.J., Arnold, S., Challinor, A.J. & Emberson, L., 2012, **Intercontinental Trans-Boundary Contributions to Ozone-Induced Crop Yield Losses in the Northern Hemisphere**, Biogeosciences, 9(1), PP. 271-292.  
<https://doi.org/10.5194/bg-9-271-2012>, 2012.
- Huang, Y., Yang, Z. & Gao, Z., 2019, **Contributions of Indoor and Outdoor Sources to Ozone in Residential Buildings in Nanjing**, International Journal of Environmental Research and Public Health, 16(14), P. 2587.  
<https://doi.org/10.3390/ijerph16142587>
- Ito, K., De Leon, S.F. & Lippmann, M., 2005, **Associations between Ozone and Daily Mortality: Analysis and Meta-Analysis**, Epidemiology, 16(4), PP. 446-457.  
<https://doi.org/10.1097/01.ede.0000165821.90114.7f>
- Javanbakht Amiri, S. & Khatami, S.H., 2012, **The Study of the Correlation between Air Quality Index Pollution and Meteorological Parameters in Tehran with Regression Analysis Approach**, Human and Environment, 10(34), PP. 15-28.
- Kampa, M. & Castanas, E., 2008, **Human Health Effects of Air Pollution**, Environmental Pollution, 151(2), PP. 362-367.  
<https://doi.org/10.1016/j.envpol.2007.06.012>
- Li, X. & Rappenglück, B., 2014, **A WRF-CMAQ Study on Spring Time Vertical Ozone Structure in Southeast Texas**, Atmospheric Environment, 97, PP. 363-385.  
<https://doi.org/10.1016/j.atmosenv.2014.08.036>
- Moustris, K., Nastos, P., Larissi, I. & Paliatsos, A., 2012, **Application of Multiple Linear Regression Models and Artificial Neural Networks on the Surface Ozone Forecast in the Greater Athens Area, Greece**, Advances in Meteorology, 2012.  
<https://doi.org/10.1155/2012/894714>
- Özbay, B., 2012, **Modeling the Effects of Meteorological Factors on SO<sub>2</sub> and PM<sub>10</sub> Concentrations with Statistical Approaches**, Clean-Soil, Air, Water, 40(6), PP. 571-577.  
<https://doi.org/10.1002/clen.201100356>
- Pak, U., Kim, C., Ryu, U., Sok, K. & Pak, S., 2018, **A Hybrid Model Based on Convolutional Neural Networks and Long Short-Term Memory for Ozone Concentration Prediction**, Air Quality, Atmosphere & Health, 11(8), PP. 883-895.  
<https://doi.org/10.1007/s11869-018-0585-1>
- Reeves, C.E., Penkett, S.A., Bauguette, S., Law, K.S., Evans, M.J., Bandy, B.J., Monks, P.S., Edwards, G.D., Phillips, G. & Barjat, H., 2002, **Potential for Photochemical Ozone Formation in the Troposphere over the North Atlantic as Derived from Aircraft Observations during ACSOE**, Journal of Geophysical Research: Atmospheres, 107(D23), ACH 14-11-ACH 14-14.  
<https://doi.org/10.1029/2002JD002415>

- Ren, X., Mi, Z. & Georgopoulos, P.G., 2020, **Comparison of Machine Learning and Land Use Regression for fine Scale Spatiotemporal Estimation of Ambient Air Pollution: Modeling Ozone Concentrations across the Contiguous United States**, *Environment International*, 142, P. 105827.  
<https://doi.org/10.1016/j.envint.2020.105827>
- Shan, W., Yin, Y., Lu, H. & Liang, S., 2009, **A Meteorological Analysis of Ozone Episodes Using HYSPLIT Model and Surface Data**, *Atmospheric Research*, 4(93), PP. 767-776.  
 DOI: 10.1016/j.atmosres.2009.03.007
- Sotoudehian, S. & Arhami, M., 2014, **Estimating Ground-Level PM 10 Using Satellite Remote Sensing and Ground-Based Meteorological Measurements over Tehran**, *Journal of Environmental Health Science and Engineering*, 12(1), PP. 1-13.  
<https://doi.org/10.1186/s40201-014-0122-6>
- Sousa, S., Martins, F.G., Alvim-Ferraz, M. & Pereira, M.C., 2007, **Multiple Linear Regression and Artificial Neural Networks Based on Principal Components to Predict Ozone Concentrations**, *Environmental Modelling & Software*, 22(1), PP. 97-103.  
<https://doi.org/10.1016/j.envsoft.2005.12.002>
- Spellman, G., 1999, **An Application of Artificial Neural Networks to the Prediction of Surface Ozone Concentrations in the United Kingdom**, *Applied Geography*, 19(2), PP. 123-136.  
[https://doi.org/10.1016/S0143-6228\(98\)00039-3](https://doi.org/10.1016/S0143-6228(98)00039-3)
- Sullivan, J.T., Rabenhorst, S.D., Dreessen, J., McGee, T.J., Delgado, R., Twigg, L. & Sumnicht, G., 2017, **Lidar Observations Revealing Transport of O3 in the Presence of a Nocturnal Low-Level Jet: Regional Implications for "Next-Day" Pollution**, *Atmospheric Environment*, 158, PP. 160-171.  
<https://doi.org/10.1016/j.atmosenv.2017.03.039>
- Susaya, J., Kim, K.-H., Shon, Z.-H. & Brown, R.J., 2013, **Demonstration of Long-Term Increases in Tropospheric O3 Levels: Causes and Potential Impacts**, *Chemosphere*, 92(11), PP. 1520-1528.  
<https://doi.org/10.1016/j.chemosphere.2013.04.017>
- Ul-Saufie, A.Z., Yahaya, A.S., Ramli, N.A., Rosaida, N. & Hamid, H.A., 2013, **Future Daily PM10 Concentrations Prediction by Combining Regression Models and Feedforward Backpropagation Models with Principle Component Analysis (PCA)**, *Atmospheric Environment*, 77, PP. 621-630.  
<https://doi.org/10.1016/j.atmosenv.2013.05.017>
- Vingarzan, R., 2004, **A Review of Surface Ozone Background Levels and Trends**, *Atmospheric Environment*, 38(21), PP. 3431-3442.  
<https://doi.org/10.1016/j.atmosenv.2004.03.030>
- Xiang, S., Liu, J., Tao, W., Yi, K., Xu, J., Hu, X., Liu, H., Wang, Y., Zhang, Y., Yang, H., Hu, J., Wan, Y., Wang, X., Ma, J., Wang, X., Tao, S., 2020, **Control of both PM2.5 and O3 in Beijing-Tianjin-Hebei and the surrounding areas**, *Atmospheric Environment*, 224, 117259.  
<https://doi.org/10.1016/j.atmosenv.2020.117259>



# نخستین از دور و GIS ایران



سنجش از دور و GIS ایران سال پانزدهم، شماره چهارم، زمستان ۱۴۰۲  
Vol.15, No. 4, Winter 2023 Iranian Remote Sensing & GIS

17-30

## Improving the Accuracy of Ground Surface Ozone Concentration Estimation Using Satellite Products and Machine Learning

Atashi Deligani R.<sup>1</sup>, Moradizadeh M.<sup>2\*</sup>, Tashayo B.<sup>2</sup>

1. M.Sc. Student, Dep. of Geomatics, Faculty of Civil and Transportation Engineering, University of Isfahan, Isfahan
2. Assistant Prof., Dep. of Geomatics, Faculty of Civil and Transportation Engineering, University of Isfahan, Isfahan

### Abstract

Ground surface ozone is one of the most dangerous pollutants that has significant harmful effects on the residents of urban areas. The purpose of this study is to identify the factors affecting ozone concentration and modeling its changes using satellite data and different machine learning methods in Tehran. For this purpose, pollutant concentration and meteorological data were used along with the satellite product of land surface temperature (LST) in the period from 2015 to 2021. After calculating the correlation between ozone concentration and independent parameters, ozone concentration modeling was done in five different modes in terms of input parameters and learning method and applying data refinement. In the first and second mode, modeling was done using pollutant concentration and meteorological data through multivariate linear regression method. The only difference between these two modes is the filtering of the input data using the WTEST method in the second mode. In the third mode, the LST product was added to the input data, and in the fourth and fifth mode, ozone modeling was done using multilayer neural network and recurrent neural network, respectively. The comparison of the five modes showed that the modeling of the first to fifth stages with adjusted coefficient of determination of 0.5, 0.64, 0.69, 0.74 and 0.8 were able to recover the ozone concentration, respectively. It was also found that among different pollutants, nitrogen monoxide, nitrogen dioxide and nitrox have the greatest impact on ozone concentration, just as temperature, humidity and wind speed are the most influential among meteorological data. Although the use of WTEST statistics led to the identification and elimination of inconsistencies and errors in the observations of pollution measurement stations, the neural network learning method showed better performance in modeling than multivariate regression due to its less sensitivity to noise. As a notable result, adding the LST product to the input data brought a 5% increase in accuracy in estimating ozone concentration.

**Keywords:** Ozone concentration, Machine learning, Multivariate linear regression, Recurrent neural network, Atmospheric pollutant.

\* Correspondence Address: Dep. of Geomatics, Faculty of Civil and Transportation Engineering, University of Isfahan, Isfahan.  
Tel: 09133063994

Email: m.moradizadeh@eng.ui.ac.ir

<https://orcid.org/0000-0003-3163-5853>

<https://dx.doi.org/10.48308/gisj.2022.102758>



Copyright: © 2023 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).