نشریه علمی- پژوهشی علوم و فنون نقشه برداری، دوره پنجم، شماره ۳، بهمن ماه ۱۳۹۴

پهنه‌بندی آلودگی ذرات معلق با استفاده از مدل‌های آماری محلی در GIS (مطالعه موردی، شهر تهران)

روح الامین بهاری [۱]، رحیم علی‌عباسپور*[۲]، پرهام پهلوانی [۲]

[۱] دانشجوی کارشناسی ارشد سیستم‌های اطلاعات مکانی- دانشکده مهندسی نقشه برداری و اطلاعات مکانی- پردیس دانشکده‌های فنی- دانشگاه تهران
bahari.giseng@gmail.com

[۲] استادیار دانشکده مهندسی نقشه‌برداری و اطلاعات مکانی- پردیس دانشکده‌های فنی- دانشگاه تهران
abaspour@ut.ac.ir.com
pahlavani@ut.ac.ir

(تاریخ دریافت مرداد ۱۳۹۴، تاریخ تصویب آذر ۱۳۹۴)

چکیده

امروزه آلودگی هوا در شهرهای بزرگ دنیا سلامت انسان‌ها را به مخاطره کشانده است. شهر تهران نیز از این قضیه مستثنی نبوده است و آلودگی در این شهر به یک چالش بزرگ برای مدیریت شهری تبدیل شده است. در پژوهش کنونی به مدلسازی مکانی غلظت آلاینده PM2.5 با بهره‌گیری از مدل رگرسیون وزندار جغرافیایی از خانواده مدلهای آماری محلی پرداخته شده است. برای این مدلسازی از ۱۸ ایستگاه سنجش کیفیت هوای شهر تهران استفاده شده و پارامترهای غلظت آلاینده PM2.5، کاربری‌های ۹ گانه، جمعیت، پارامترهای هواشناسی، فاصله از معابر اصلی و آزادراه‌ها و ارتفاع به عنوان پارامترهای مؤثر در نظر گرفته شده‌اند. از ویژگیهای این مدل ترکیب متغیرهای مختلف با همبستگی فضایی متنوع می‌باشد. این پژوهش در نهایت منجر به تولید نقشه‌هایی از وضعیت آلاینده PM2.5 بر روی کل شهر تهران شده است که به منظور شناخت مناطق پر ریسک در شهر و بکارگیری اقدامات مفید به‌منظور کاهش آلودگی در آن نقاط بسیار سودمند می‌باشد. پس از مقایسه این نقشه‌ها با مقادیر مشاهده شده و بررسی پارامترهای آماری ضریب تعیین (R^2=0.75 - 0.80) و جذر میانگین مربعات خطا (RMSE=7.1 - 8.5) مشخص گردید که مدل پیشنهادی توانایی بالایی در تخمین غلظت در مناطق مختلف در سطح شهر تهران دارد.

واژگان کلیدی: آلودگی هوا، پهنه‌بندی، PM2.5، مدلهای آماری محلی، رگرسیون وزندار جغرافیایی، R^2

* نویسنده رابط

Archive of SID

www.SID.ir

۱- مقدمه

امروزه آلاینده‌های گوناگونی به خاطر اقدامات انسانی و فعالیت های بیولوژیکی وارد هوا شده و زندگی انسان‌ها را به مخاطره می‌کشاند؛ چرا که این آلاینده‌ها اثر مستقیم بر سلامت انسان‌ها دارند. تهران به عنوان پایتخت ایران نیز از این امر مستثنی نبوده و دائماً با این خطرات دست به گریبان می‌باشد. در سال‌های اخیر یکی از آلاینده هایی که بیشترین تهدید را برای شهر تهران داشته است ذرات معلق بوده است که در این میان ذرات معلق با قطر کمتر از ۲/۵ میکرون دارای بیشترین روزهای ناسالم در طی سال‌های اخیر بوده است [۷]. این ذرات ممکن است با منشأ طبیعی (گردگیاهان، هاگ‌ها، باکتری‌ها، ویروس‌ها، تک یاخته‌ای‌ها، قارچ‌ها، رشته‌های گیاهی ، بقایای زنگ زدگی و غبار ناشی از فعالیت های آتشفشانی) و یا انسانی (فعالیت های احتراقی دود ناشی از خاکستر، دود، دوده ها، اکسیدهای فلزی، نمک ها، قطرات روغنی یا قیری، قطرات اسیدی، سیلیکات‌ها، سایر غبارهای معدنی و دودهای غلیظ فلزی) باشد. مطالعات بهداشتی، ارتباط معناداری بین قرارگرفتن در معرض ریزگردها و مرگ زودرس ناشی از بیماری‌های قلبی و ریوی نشان داده اند [۶].

از اینرو، افزایش غلظت آلاینده‌ها به یک چالش بزرگ برای مدیریت کلانشهرهایی مانند تهران تبدیل شده است. اطلاع از وضعیت پراکندگی مکانی آلودگی آلاینده‌ها در سطح شهر این امکان را به مدیران کلانشهرها می‌دهد تا با در نظر گرفتن اقداماتی مناسب میزان خطرات را برای مناطق و افراد ریسک‌پذیر کاهش دهند. از طرفی اندازه گیری غلظت آلاینده ها هزینه بر بوده و بصورت نقطه‌ای انجام می‌گیرد. اما ضرورت ایجاب می‌کند این اطلاعات اندازه‌گیری شده در ایستگاه‌ها برای تحلیل‌های منطقه‌ای به کل منطقه مورد مطالعه تعمیم داده شوند و توزیع آنها در پهنه شهر ترسیم شود.

به طور کلی دو روش کلی درونیاب‌های مکانی و مدل‌های پراکندگی به منظور پهنه‌بندی آلاینده‌ها وجود دارد. در سال های اخیر با توسعه روش‌های آماری و مدل‌های زمین آمار، مدل‌های درونیابی مکانی متعددی در دسترس محققان قرار گرفته است. روش‌های درونیابی از نقاطی با مقادیر معلوم در اطراف نقاط با مقادیر مجهول استفاده کرده و مقادیر آنها را تخمین می‌زنند[۱۴]، اما چگونگی تأثیر نقاط معلوم در تخمین مقادیر مجهول سبب می‌شود تا روش‌های درونیابی به دو دسته سراسری و محلی تقسیم‌بندی شوند. در دسته اول، یک صفحه به کل داده‌ها برازش داده می‌شود و در دسته دوم بر اساس نقاط نزدیک درونیابی صورت می‌گیرد.

در مطالعات مکانی با داده هایی روبرو هستیم که از نظر مکانی متغیر هستند و با حرکت در منطقه مورد مطالعه از نقطه-ای به نقطه دیگر مقادیر تغییر می‌کند. از طرفی در مدلسازی-هایی که پارامتر مورد مطالعه تحت تأثیر متغیرهای مستقلی است که در جای جای منطقه تغییر می‌کنند، در چنین مواردی اگر از روشهای سراسری استفاده شود، در نهایت در ماتریس وزن یا وابستگی نهایی برای هر متغیر مستقل، یک مقدار دیده می‌شود که این به این معناست که در روش سراسری برای کل منطقه ارتباط پارامتر مورد مطالعه با هر متغیر مستقل را یکی در نظر می‌گیرند که در بسیاری از موارد این امر با واقعیت متفاوت بوده و وابستگی با تغییر مکان و جابجایی تغییر می‌کند. در روش‌های محلی برخلاف روش‌های سراسری در اطراف هر نمونه بر روی منطقه مورد مطالعه محدوده‌ای در نظر گرفته می‌شود و وزن و وابستگی بین متغیر وابسته و متغیر یا متغیرهای مستقل محاسبه می‌گردد. در اینصورت دیگر وزن‌ها و ضرایب وابستگی، ثابت نبوده و بصورت محلی تغییر می‌کنند. در این صورت مشاهداتی که به هم نزدیکترند نسبت به آن‌هایی که از هم دورترند، منعکس کننده وابستگی مکانی بالاتر هستند.

در این پژوهش جهت پهنه‌بندی آلاینده PM2.5 بر روی شهر تهران از روش رگرسیون وزن‌دار جغرافیایی که یکی از روش‌های رگرسیونی محلی است استفاده شده است. در این پژوهش از پارامترهای کاربری، جمعیت، ارتفاع، راه‌های اصلی و آزادراه‌ها، دما و سرعت و جهت باد و غلظت آلاینده جهت ورود به مدل استفاده شده است. روند کلی به این صورت است که ابتدا به کمک غلظت و پارامترهای بیان شده و مدل رگرسیون وزندار جغرافیایی ماتریس وزن وپارامترهای مؤثر محاسبه می‌شود و سپس این وزن‌ها به ۲۱۰۰ نقطه همراه با مقادیر مؤثر مذکور اعمال می‌شود و غلظت این نقاط استخراج می‌گردد که در نهایت به کمک مدل کریجینگ و غلظت تخمین زده شده برای این نقاط یک سطح که بیانگر غلظت PM2.5 بر روی شهر تهران است استخراج می‌گردد که در قسمت‌های بعدی توضیح داده خواهد شد.

۲- پیشینه تحقیق

در سالهای اخیر مطالعات متعددی در زمینه پهنه‌بندی در داخل و خارج از کشور صورت گرفته است Kavousi (۲۰۱۳) و

Archive of SID

www.SID.ir

همکاران به‌منظور پهنه‌بندی آلاینده‌های PM10 و CO بر روی شهر تهران و شناسایی مکان‌های آسیب پذیر برای سال ۱۳۹۰ از روش کریجینگ استفاده کرده‌اند. نتایج نشان داد که خروجی‌های حاصل از مدل در مرکز شهر دارای دقت بالاتری می‌باشند.

Robinson (۲۰۱۳) و همکاران برای پهنه‌بندی آلاینده NO2 از سه روش simple kriging، ordinary kriging و کریجینگ ساده همراه با میانگین متغیرهای محلی استفاده نموده‌اند که دو روش اول تنها از متغیر اولیه به عنوان داده استفاده می‌کنند، ولی در روش سوم از داده‌های دیگر نیز استفاده می‌شود. فتح تبار(۱۳۹۰) و همکاران به‌منظور پهنه‌بندی آلاینده‌های هوا از از داده‌های سال ۱۳۸۸ و دو روش کریجینگ ساده به عنوان یک روش زمین‌آماری در حالت‌های مختلف واریوگرام و روش اسپیلاین در حالتهای مختلف در فصول سال استفاده و ارزیابی گردیده است. برای ارزیابی دقت نیز از دو شاخص RMSE و MAE استفاده شده است. روش کریجینگ در حالتهای مختلف با واریوگرام با سطح خطای کمتر به عنوان روش بهینه انتخاب شده است.

رفیع‌پور(۱۳۹۲) و همکاران در پژوهشی به‌منظور مدلسازی مکانی آلاینده منوکسیدکربن از سیستم اطلاعات مکانی، رگرسیون چند متغیره و شبکه‌های عصبی استفاده کردند. جهت نیل به این منظور آنها از ۱۸ ایستگاه سنجش کیفیت هوا و داده‌های موقعیت، ارتفاع و فاصله از خیابان اصلی جهت ورود به مدل استفاده نموده‌اند. آنها جهت مدلسازی از رگرسیون چند متغیره و سه شبکه عصبی MLP، RBF و GRNN استفاده نمودند که در نهایت نتایج نشان داد که شبکه عصبی MLP کارایی بهتری نسبت به بقیه مدل‌ها دارد.

کاوسی(۱۳۹۲) و همکاران در مطالعه‌ای به منظور پهنه-بندی غلظت آلاینده CO و تهیه نقشه احتمال رخداد آلودگی هوا برای این آلاینده، از سه مدل اتولجستیک، اتولجستیک مرکزی شده و روش کریجینگ برای شهر تهران استفاده کرده‌اند. در این پژوهش از داده‌های غلظت CO ایستگاه‌های سنجش کیفیت هوای شهر تهران برای دیماه ۱۳۹۰ استفاده شده است. نتایج حاصل نشان داد که مناطق شمالی، شرقی و مرکزی شهر تهران از سایر نواحی شهر آلوده تر هستند و از بین سه روش مذکور، مدل اتولجستیک مرکزی از توان بالاتری نسبت به روش کریجینگ و مدل اتولجستیک برخوردار است. در این پژوهش پیشنهاد شده است که برای تحلیل داده‌های دودویی از مدل اتولجستیک مرکزی استفاده گردد، زیرا در این مدل با لحاظ شدن اثر متغیرهای توضیحی، پیشگویی‌های دقیق‌تری نسبت به روش کریجینگ ارائه می‌گردد.

Kumar (۲۰۱۱) و همکاران بمنظور مدلسازی مکانی سه آلاینده SPM، SO2 و NO2 برای محیط بندر بلر در هند دو مدل معکوس فاصله وزندار و کریجینگ را مورد آزمون قرار دادند. برای مقایسه مقادیر درونیابی شده از سه پارامتر MAE، RMSE و d استفاده کرده‌اند. داده‌های مورد استفاده برای پارامترها مورد تایید مرکز کنترل آلودگی هند بوده است. پس از مقایسه مقادیر اندازه‌گیری شده و درونیابی شده مشخص گردید که مدل IDW به نسبت کریجینگ از تشابه بیشتری برای هر سه آلاینده برخوردار است. برای مدل IDW به نسبت کریجینگ مقادیر MAE و RMSE پایین و مقدار d بالاتر است که نشان دهنده تشابه بهتر بین مقادیر مشاهده شده و تخمین زده شده می‌باشد.

بررسی‌ها نشان می‌دهد که در بیشتر پژوهش‌ها تأثیر پارامترهای مؤثر دیده نشده است و تنها قابلیت روشها با خود پارامتر مورد نظر ارزیابی شده است. در برخی دیگر از مطالعات نیز تأثیر پارامترهای مؤثر در نظر گرفته شده، ولی این اثر به‌صورت سراسری و در همه نقاط دامنه مورد مطالعه به‌صورت ثابت در نظر گرفته شده است که در این مطالعه، علاوه بر لحاظ کردن اثر پارامترهای مؤثر، این اثر در همه جای دامنه مورد مطالعه تغییر می‌کند، به عبارت دیگر، مدل به‌صورت محلی اقدام به تخمین غلظت می‌نماید.

۳- منطقه مورد مطالعه

کلانشهر تهران، پایتخت ۲۲۰ ساله ایران در کوهپایه‌های جنوبی رشته کوه البرز در حد فاصل طول ۵۱ درجه و ۵ دقیقه‌ی شرقی تا ۵۱ درجه و ۵۳ دقیقه‌ی شرقی، و عرض جغرافیایی ۳۵ درجه و ۳۴ دقیقه‌ی شمالی تا ۳۵ درجه و ۵۹ دقیقه‌ی شمالی با حدود ۷۰۰ کیلومتر مربع مساحت گسترده شده است. این شهر از شمال به سلسله جبال البرز، از شرق به لواسانات و از غرب به کرج و از جنوب به ورامین محدود می‌شود. ارتفاع شهر در بلندترین نقاط شمال به ۲۰۰۰ متر و در جنوبی‌ترین نقاط به ۱۰۵۰ متر از سطح دریا می‌رسد. شهر تهران با جمعیتی قریب به ۱۲ میلیون نفر (به همراه شهرهای اقماری خود) ۱۲ درصد جمعیت کل کشور را به خود اختصاص داده است. با توجه به اینکه شهر تهران دارای موقعیت خاص جغرافیایی است (اختلاف ارتفاع زیاد در شمال

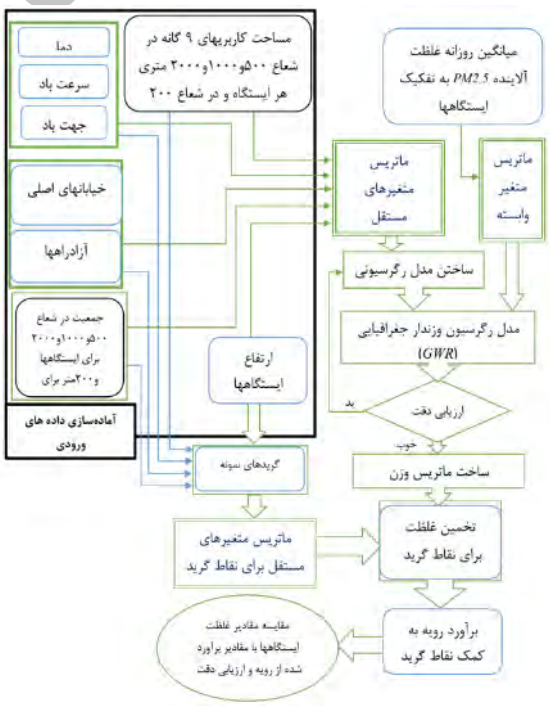
Archive of SID

www.SID.ir

و جنوب آن) و از شرایط نامناسب بافت شهری برخوردار است و وسائط نقلیه زیادی در طول شبانه روز در آن به فعالیت مشغولند. بادهای غربی در تمام طول سال دود کارخانجات و سایر عوامل آلاینده را به سطح شهر تهران وارد می‌سازند، در مجموع دارای شرایط نامساعد زیست محیطی بوده و آلودگی هوای آن در سالهای اخیر با محتوای گازهای سمی به صورت خطرناک عمل می‌نماید که تغییرات بسیار چشمگیر محیطی و اقلیمی را در آن موجب گردیده است.

۴- روش پیشنهادی

در این پژوهش هدف پهنه‌بندی غلظت آلاینده PM2.5 به کمک مقادیر ایستگاه‌های سنجش آلودگی و استفاده از پارامترهای محیطی مؤثر بر غلظت از قبیل پارامترهای هواشناسی (دما، سرعت و جهت باد)، کاربری‌های ۹ گانه، راه‌های اصلی و آزادراه‌ها، جمعیت و ارتفاع بر روی شهر تهران می‌باشد. برای دستیابی به این مهم، مدل GWR انتخاب و پیاده‌سازی شده است که در شکل (۱) الگوریتم مدل پیاده‌سازی شده به اجمال نمایش داده شده است.



شکل ۱- الگوریتم مدل پیشنهادی

همانطور که در شکل (۱) نشان داده شده است، در این مدل، در ابتدا داده‌ها در دو گروه ایستگاه‌ها و نقاط گرید آماده-سازی می‌شوند. سپس میانگین غلظت ایستگاه‌ها به همراه داده های مستقل آماده‌سازی شده در مرحله قبلی، وارد مدل رگرسیون وزندار جغرافیایی می‌شوند، پس از ارزیابی مدل‌های مختلف با چیدمان‌های متفاوت از داده‌های مستقل بهترین مدل انتخاب و ماتریس وزن آن استخراج می‌گردد. در مرحله بعد ماتریس وزن استخراج شده به همان چیدمان از داده‌های مستقل برای نقاط گرید اعمال و غلظت در آن نقاط تخمین زده می‌شود. در مرحله نهایی به کمک غلظت نقاط گرید و مدل کریجینگ رویه‌ای از غلظت بر روی شهر برازش داده می‌شود و مقدار این رویه در محل ایستگاه‌ها با مقدار غلظت واقعی ایستگاه مقایسه و مدل ارزیابی می‌گردد.

۴-۱- آماده‌سازی داده‌های مورد استفاده

به منظور مدلسازی و پهنه‌بندی آلاینده از پارامترهای متنوعی استفاده شده است که عبارتند از:

- غلظت آلاینده PM2.5 : که از میانگین روزانه غلظت PM2.5 ایستگاه‌های سنجش آلودگی سازمان محیط زیست و مرکز کنترل کیفیت هوای شهر تهران در سال ۱۳۹۲ استخراج شده است (موقعیت مکانی این ایستگاه‌ها در شکل ۲ نمایش داده شده است).
- ارتفاع ایستگاه‌های سنجش آلودگی و نقاط گرید که از مدل ارتفاعی رقومی ۳۰ متری شهر تهران (LAND SAT 7) استخراج شده است(شکل ۲).
- جمعیت در اطراف ایستگاه‌ها و نقاط گرید که برای هر ایستگاه سنجش آلودگی در شعاع‌های ۵۰۰، ۱۰۰۰ و ۲۰۰۰ متر و برای هر نقطه گرید در شعاع ۲۰۰ متر و به کمک بلوک‌های جمعیتی شهر تهران محاسبه گردیده است.
- کاربری اراضی شهر تهران که شامل ۹ کاربری (مسکونی، صنعتی، اداری و تجاری، فضای سبز، نظامی، بایر، زمین‌های مزروعی، حمل و نقل و انبارداری، انواع خدمات و تأسیسات شهری) می‌باشد. این کاربری‌ها از نقشه‌های کاربری شهرداری تهران استخراج و سپس مساحت هر کاربری برای ایستگاه‌های سنجش آلودگی در شعاع ۵۰۰ ، ۱۰۰۰ و ۲۰۰۰متر و برای هر نقطه گرید در شعاع ۲۰۰ متر محاسبه گردید.
- فاصله از راه‌های اصلی و تندراه‌ها: که در این پژوهش نقشه راه‌های شهر تهران تهیه و راه‌ها در دو دسته راه‌های اصلی و تندراه‌ها تقسیم بندی شده‌اند. سپس

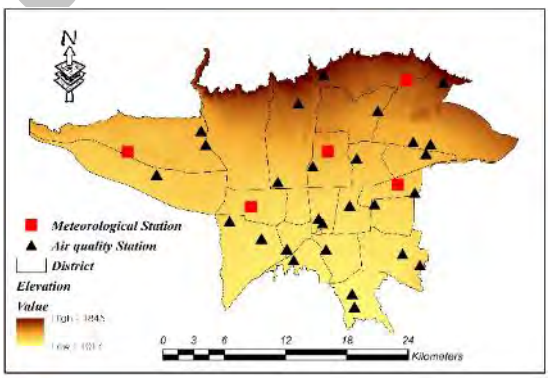
www.SID.ir

فاصله هر ایستگاه سنجش آلودگی و نقاط گرید بصورت جداگانه از این دسته از راه‌ها محاسبه گردید.

- پارامترهای هواشناسی دما، سرعت و جهت باد در هر روز: ابتدا میانگین این پارامترها برای پنج ایستگاه هواشناسی سینوپتیک مستقر در شهر تهران (شکل۲) محاسبه و سپس به کمک روش کریجینگ بر روی کل شهر تعمیم داده شد. پس از بدست آمدن لایه‌ها بر روی شهر تهران مقادیر این سه پارامتر برای ایستگاه‌ها و نقاط گرید از این سطوح استخراج و وارد مدل گردید.

در این پژوهش به این دلیل که اطلاعات ورودی هر کدام دارای یکای خاص خود می‌باشند و دامنه مقادیر با هم متفاوت هستند وزن‌های با مقدار عددی خیلی بزرگ یا خیلی کوچک در طول محاسبات ایجاد می‌گردد. به منظور جلوگیری از این رخداد داده‌ها نرمالسازی شدند. در این مقاله داده‌ها با استفاده از رابطه (۱) در محدوده [0,1] نرمال سازی شده‌اند.

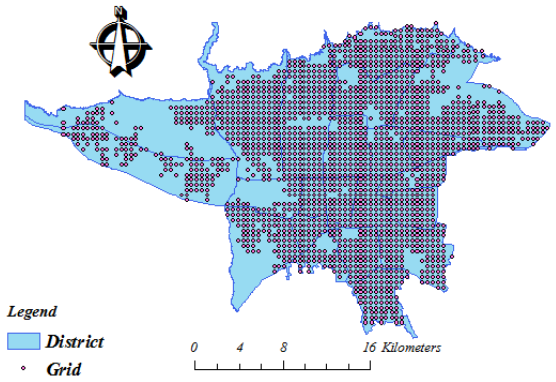
$$X_{nor} = \left(\frac{x - x_{min}}{x_{max} - x_{min}}\right) \qquad (۱)$$



شکل ۲- پراکندگی ایستگاه‌های سنجش آلودگی هوا، ایستگاه‌های هواشناسی و نقشه ارتفاعات تهران

۴-۲- نقاط گرید

در این مدل ۲۱۰۰ نقطه بر روی شهر تهران بصورت یک گرید منظم اما تصادفی، بطوریکه منطقه‌های تحت پوشش هر نقطه با مناطق تحت پوشش نقاط کناری تقاطع نداشته باشند و سپس همه پارامترهای مستقلی که در بخش قبل ذکر گردید، برای این نقاط محاسبه شد (در شعاع ۲۰۰ متر برای کاربری‌ها و جمعیت). در نهایت از این مقادیر ماتریسی تهیه شد که تعداد سطرهایش برابر تعداد نقاط و تعداد ستون آن برابر با تعداد پارامترهای مستقل می‌باشد. شکل ۳ نشان دهنده موقعیت این نقاط می‌باشد.



شکل ۳- موقعیت ۲۱۰۰ نقطه گرید پراکنده شده بر روی شهر تهران

۴-۳- ساخت مدل رگرسیون

یکی از مهمترین مراحل در بسیاری از مسائل رگرسیون چندگانه (که با تعداد زیادی متغیر مستقل مواجه هستند) انتخاب زیر مجموعه‌ای از مناسب‌ترین متغیرها برای پیش‌بینی متغیر وابسته است. در واقع، با استفاده از یک دسته متغیر مستقل مشخص می‌توان مدل‌های متفاوتی ساخت. برای مثال با ۵ متغیر مستقل می‌توان ۳۲ مدل مختلف ایجاد کرد. هرگاه تعداد متغیرهای مستقل کم باشد، می‌توان تمامی مدل‌های ممکن را بررسی نمود، ولی روش‌های مختلفی وجود دارد که به محاسبات متعددی نیاز ندارند. در اینجا سه روش اصلی که در این تحقیق استفاده شده است، تشریح می‌گردد.

۴-۳-۱- انتخاب رو به جلو[۱]

در این روش در مرحله اول، مدل تنها مقدار ثابت را دارد. در هر مرحله، متغیری به مدل اضافه می‌شود که بیشترین تغییر را در پارامتر ضریب تعیین (در بخش ۴-۵ توضیح داده خواهد شد) R^2 ایجاد کند و این تغییر در R^2 باید به حدی باشد که بتواند این فرض را که مقدار واقعی تغییر برابر با صفر است، رد کند. این مدل تا جایی متغیر وارد می‌کند که متغیر دیگری وجود نداشته باشد که بتواند افزایش معناداری در R^2 ایجاد کند.

۱ Forward selection

Archive of SID

www.SID.ir

۴-۳-۲- حذف رو به عقب[۱]

در این روش ابتدا تمامی متغیرها در مدل وجود دارند. در هر مرحله متغیری از مدل خارج می‌شود که کمترین تغییر را در مقدار R^2 ایجاد کند و این تغییر در مقدار R^2 باید به حدی باشد که نتوان این فرضیه را رد کرد که مقدار واقعی تغییر R^2 برابر با صفر است. خارج کردن متغیرها از مدل هنگامی متوقف می‌شود که خارج کردن هر کدام از متغیرها از مدل، تغییر معناداری در R^2 ایجاد کند.

۴-۳-۳- انتخاب قدم به قدم[۲]

وارد شدن متغیرها به مدل در این روش شبیه به روش رو به جلو است با این تفاوت که هر بار بعد از وارد کردن یک متغیر به مدل، تمام متغیرهایی که تاکنون وارد مدل شده‌اند و پیش‌بینی کننده معناداری نیستند، از مدل خارج می‌گردند. در واقع این روش ترکیبی از دو روش قبلی است. برای این کار ابتدا دو متغیر اول به روش رو به جلو وارد مدل می‌شود و سپس به روش رو به عقب این دو را بررسی می‌کند تا به این نتیجه برسد که آیا یکی از این دو کاندیدای خروج از مدل می‌باشند یا خیر. اگر چنین باشد از مدل حذف می‌شوند. و در هر مرحله با همان قواعدی که روش رو به جلو دارد، متغیری را وارد مدل می‌کند و سپس با استفاده از قواعد حذف رو به عقب کلیه متغیرهایی که تاکنون وارد مدل شده‌اند را بررسی کرده و اگر معیار لازم را داشته باشند، از مدل حذف می‌شوند. این فرآیند تا بررسی تمام متغیرها ادامه می‌یابد.

۴-۴- روش رگرسیون وزندار جغرافیایی

در سال ۱۹۹۸، برانسدون و همکارانش راه حلی را پیشنهاد کردند که رگرسیون وزندار جغرافیایی[۳] *(GWR)* نام گرفت[۸]. این روش برخلاف رگرسیون کمترین مربعات معمولی، ویژگی خودهمبستگی داده‌ها را در نظر گرفته و بر خلاف مدل‌های رگرسیون مکانی سراسری، یک مجموعه پارامترهای برآورد شده از نظر مکانی متغیر[۴] را ارائه می‌کند که شامل ضرایب خودهمبستگی مکاناً متغیر می‌باشند. بنابراین رگرسیون وزندار جغرافیایی در مقایسه با رگرسیون‌های مکانی، یک روش جایگزین و ساده برای استخراج معیارهای محلی خود همبستگی می‌باشد. مزیت عمده *GWR* در مقابل الگوهای رگرسیون معمولی، توانایی آن در امکان بررسی ناایستایی مکانی است[۹]. ناایستایی مکانی نشان می‌دهد که اندازه‌گیری یا تخمین روابط بین متغیرها از محلی به محل دیگر تفاوت می‌کند. علاوه براین، روش *GWR* یک رگرسیون محلی و موضعی است که به طور معناداری رگرسیون معمولی را برای استفاده در داده-های مکانی بهبود داده است. برخلاف الگوهای رگرسیون معمولی که یک معادله را برای توصیف روابط کلی بین متغیرها برقرار می‌کنند، GWR اطلاعات مکانی‌ای تولید می‌کند که تغییرات مکانی بین روابط متغیرها را بیان می-کند. همچنین، *GWR* ابزاری مفید و عملی برای ارزیابی غیریکنواختی یک متغیر وابسته است. غیر یکنواختی مکانی هنگامی که ساختار فرآیند الگوسازی شده تغییر می‌کند بوجود می‌آید. *GWR* معادلات جداگانه‌ای را با مشارکت متغیرهای مستقل و وابسته‌ای که در داخل یک نوار (فاصله) از هر پدیده قرار می‌گیرند تشکیل می‌دهد. شکل و اندازه پهنای باند برای نوع کرنل، روش پهنای باند، فاصله و تعداد پدیده‌ها توسط کاربر به عنوان ورودی در نظر گرفته می-شود. *GWR* یک معادله تولید می‌کند، بنابراین این امکان را می‌دهد تا مقادیر پارامتر بصورت پیوسته در فضای جغرافیایی تغییر کند. هر یک از معادلات با استفاده از یک وزن متفاوت از مشاهدات مشتمل در مجموع داده‌ها بدست می‌آید. چون معادله رگرسیونی به طور غیر مستقل برای هر مشاهده برازش داده شده است، لذا یک پارامتر پیش‌بین جداگانه، مقدار Z و $R2$ را برای هر مشاهده محاسبه می-کنند. جزئیات کامل *GWR* توسط فادرینگهام و همکاران ارائه شده است[۹].

تحلیل‌های رگرسیونی برای الگوبندی رابطه بین یک متغیر با یک یا چند متغیر دیگر متغیر دیگر مطابق روابط ۲ و۳ استفاده می‌شود.

$$y_i = \beta_{0i} + \beta_{1i}X_{1i} + \beta_{2i}X_{2i} + \cdots + \beta_{kn}X_{kn} + \varepsilon_i \qquad (۲)$$

$$y_i = \beta_0 + \sum_k \beta_k X_{ik} + \varepsilon_i \qquad (۳)$$

---

۱ Backward elimination
۲ Stepwise
۳ Geographically Weighted Regression (GWR)
۴ Parameter estimates Spatially varying

Archive of SID

www.SID.ir

در این روابط، y_i مقدار درونیابی شده در موقعیت i، مقدار β_0 عرض از مبدأ، β_{ik} برابر است با K امین پارامتر موضعی در i امین موقعیت، X_{ik} نشان دهنده K امین متغیر مستقل در i امین موقعیت و n بیانگر موقعیت قبلی است. وزن اختصاص داده شده به هر یک از مشاهدات در GWR بر اساس یک تابع تنزل فاصله در مرکز مشاهده i است. الگوی رگرسیون وزنی جغرافیایی GWR موقعیت مکانی نمونه‌ها را در نظر می‌گیرد و این امکان را می‌دهد تا پارامترهای تخمین زده شده به‌صورت موضعی تغییر کند. یک الگو GWR می‌تواند بصورت رابطه ۴ نوشته شود:

$$y_i = \beta_0(u_i, v_i) + \sum_k \beta_k(u_i, v_i) X_{ik} + \varepsilon_i \qquad (۴)$$

در این رابطه بردار (u_i, v_i) مختصات i امین موقعیت را نشان می‌دهد و $\beta_0(u_i, v_i)$ و $\beta_k(u_i, v_i)$ پارامترهای تخمین زده شده برای i امین موقعیت هستند که مقادیر آنها با موقعیت تغییر می‌کند. X_{ik} و ε_i به‌ترتیب متغیرهای مستقل و میزان خطا در موقعیت i می‌باشند.
پارامترهای الگوی رگرسیون چندمتغیره خطی بر اساس حداقل مربعات معمولی به‌صورت ماتریس رابطه ۵ تخمین زده می‌شود:

$$\hat{\beta} = (X^T X)^{-1} X^T Y \qquad (۵)$$

در این رابطه X ماتریس مشاهدات و Y بردار پاسخ است. پارامترها در GWR با استفاده از یک تابع وزنی به‌صورت رابطه ۶ تخمین زده می‌شود:

$$\hat{\beta} = (X^T W(u_i, v_i) X)^{-1} X^T W(u_i, v_i) Y \qquad (۶)$$

در این رابطه $W(u_i, v_i)$ وزن‌هایی برای نقطه iام می‌باشند به طوری که نقاطی که به نقطه مورد مطالعه نزدیکتر هستند نسبت به نقاط دورتر تأثیر بیشتری بر نتیجه داشته باشند[۹].

۵-۴- معیارهای ارزیابی مدل

در این پژوهش از سه پارامتر آماری جهت آنالیز عملکرد مدل استفاده گردید. نخست پارامتر RMSE (رابطه ۷) که بیانگر مجموع باقیمانده‌ها است که از اختلاف مقدار واقعی غلظت و مقدار پیش بینی شده بدست می‌آید و از طرفی بیان کننده قدرت پیش بینی مدل می‌باشد .پارامتر دیگر R^2 (رابطه ۸) است که بیانگر مناسب بودن برازش می‌باشد. مقادیر این پارامتر بین صفر تا یک بوده و هر چه به یک نزدیک تر باشد نشان می‌دهد که برازش ما برازش خوبی بوده است و مدل به خوبی برازش داده شده است. پارامتر سوم RS (رابطه ۹) می‌باشد که برابر مجموع مربع باقیمانده‌ها بوده و هر چه مقدار این پارامتر کمتر باشد نشانه نزدیکتر بودن برازش به مقدار واقعی است.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(X_{obs,i} - X_{model,i})^2}{n}} \qquad (۷)$$

$$R^2 = \left(\frac{\sum_{i=1}^{n}(X_{model,i} - \bar{X})(X_{obs,i} - \bar{X})}{n.\, s_{pred}.\, s_{obs}}\right)^2 \qquad (۸)$$

$$RS = \sum_{i=1}^{n}(X_{obs,i} - X_{model,i})^2 \qquad (۹)$$

که در این روابط $X_{obs,i}$ غلظت مشاهده شده در iامین ایستگاه، $X_{model,i}$ غلظت بدست آمده از مدل در iامین ایستگاه، s_{obs} انحراف معیار استاندارد برای مشاهدات غلظت و s_{pred} انحراف معیار استاندارد برای غلظت‌های تخمین زده شده توسط مدل می‌باشد.

۵- پیاده سازی و ارزیابی نتایج

در این پژوهش برای فصل زمستان در سال ۹۲ به دلیل عدم وجود یک روز مشترک به همراه داده برای همه ایستگاه‌ها، امکان پیاده‌سازی مدل وجود نداشت و مدل پیشنهادی برای سه روز در فصل‌های بهار، تابستان و پائیز پیاده‌سازی و ارزیابی شد. به این ترتیب که ابتدا در هر روز میانگین روزانه غلظت آلاینده PM2.5 برای ایستگاه‌های سنجش کیفیت هوا محاسبه و به همراه پارمترهای اشاره شده در قسمت (۳-۲) که برای هر ایستگاه محاسبه شده بود، وارد مدل رگرسیون وزندار جغرافیایی با نوع کرنل Adaptive و با روش پهنای باند Akaike Information Criterion (AICc) شده‌اند و پس از اجرای مکرر مدل با چیدمان‌های مختلف از پارامترهای مستقل، بهترین ساختار انتخاب و ماتریس وزن استخراج گردید (جدول۱).

Archive of SID

www.SID.ir

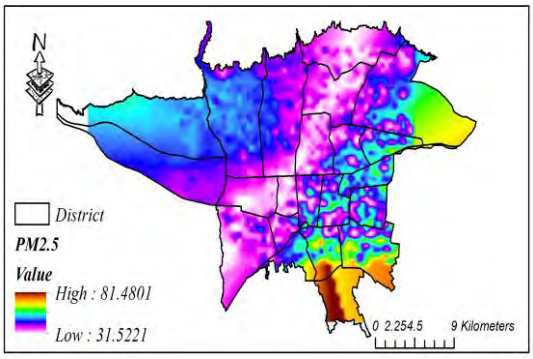
جدول ۱- پارامترهای مؤثر در هر روز و نتایج حاصل از اعمال رگرسیون وزندار جغرافیایی بر روی پارامترهای مستقل و غلظت آلاینده PM2.5

روز	بهترین ساختار	RS	R2
بهاری	مسکونی، فضای سبز، تجاری و اداری، حمل و نقل، نظامی، جمعیت، دما، فاصله از آزادراهها	۰/۲۲	۰/۷۸
تابستانی	مسکونی، فضای سبز، خدماتی، تجاری و اداری، حمل و نقل، کشاورزی، بایر، سرعت باد، ارتفاع، فاصله از خیابانهای اصلی	۰/۰۹	۰/۹۱
پائیزی	مسکونی، فضای سبز، خدماتی، حمل و نقل، بایر، جمعیت، سرعت باد، دما، ارتفاع، فاصله از آزادراهها	۰/۰۹	۰/۹۱

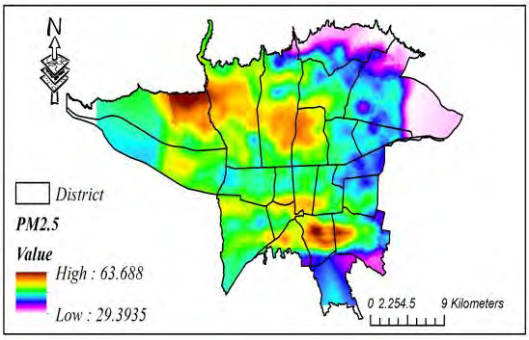
در ادامه، ماتریس وزن بدست آمده به نقاط گرید با همان ساختار بدست آمده اعمال و مقدار غلظت در آن نقاط تخمین زده شدند. در نهایت، به کمک غلظت تخمینی ۲۱۰۰ نقطه گرید و روش کریجینگ معمولی با واریوگرام نمایی و اندازه پیکسل خروجی ۱۰۸ متری سه رویه برای روزهای سه‌گانه برازش داده شد که در شکل ۴ این رویه‌ها به نمایش گذاشته شده است. پس از ساخت رویه‌ها به منظور ارزیابی دقت آنها، مقادیر مشاهده شده برای هر ایستگاه با مقادیر بدست آمده از رویه برای همان ایستگاه‌ها مقایسه و به کمک پارامترهای آماری ارزیابی گردید که نتایج آن در جدول ۲ نمایش داده شده است.

جدول ۲- مقایسه مقادیر استخراجی از رویه‌ها و مقادیر مشاهده شده در ایستگاههای سنجش کیفیت هوا برای روزهای سه‌گانه

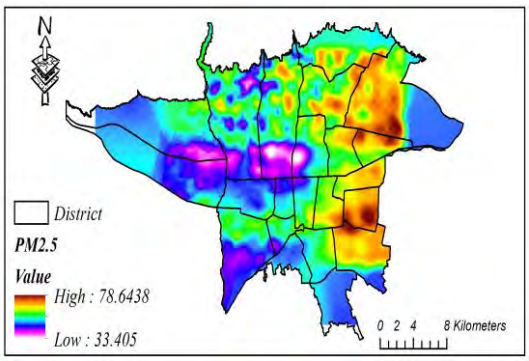
روز	RMSE	R^2
روز بهاری	۷/۴۴	۰/۷۵
روز تابستانی	۸/۵	۰/۷۹
روز پاییزی	۷/۱	۰/۸۰



شکل ۴- الف) وضعیت آلودگی بر روی شهر تهران برای روز بهاری



شکل ۴- ب) وضعیت آلودگی بر روی شهر تهران برای روز تابستانی



شکل ۴- ج) وضعیت آلودگی بر روی شهر تهران برای روز پائیزی

۶- نتیجه‌گیری

PM2.5 یکی از آلاینده‌های اصلی هوا برای شهر تهران می‌باشد و اثرات بسیار مخربی بر سلامت انسان دارد. مدلسازی مکانی آلودگی هوا می‌تواند به مدیران جهت مدیریت و شناسایی مناطق آلوده و به تبع آن بکارگیری اقدامات مناسب جهت کاهش خطر کمک می‌کند. به همین جهت در این مقاله به مدلسازی مکانی آلاینده PM2.5 با روش رگرسیون وزندار جغرافیایی پرداخته شد.

در اغلب مطالعات موجود برای مدلسازی مکانی به کمک داده‌های آماری و بکار بردن روش‌های درونیابی اقدام به تخمین و پهنه‌بندی آلاینده بر روی منطقه مورد مطالعه می‌نمایند، اما در این پژوهش سعی شد تا همزمان اثر پارمترهای مؤثر بر آلاینده در نقطه نقطه شهر در نظر گرفته شود که بدین منظور از روش رگرسیون وزندار جغرافیایی استفاده گردید.

روش رگرسیون وزندار جغرافیایی یک روش آماری است که برای مطالعه الگوهای موضعی سازگار شده است بنابر این روش مناسبی برای پهنه‌بندی و تخمین داده‌های مکانی به حساب می‌آید. نتایجی که از این تحقیق بدست آمد نیز نشان داد که روش رگرسیون وزندار جغرافیایی قدرت بالایی در تحلیل مکانی تغییرات غلظت آلاینده PM2.5 در شهر تهران

Archive of SID

www.SID.ir

دارد. پس از پهنه‌بندی و ارائه رویه‌ها نتایج حاصل از مقایسه مقادیر رویه‌ها و مقادیر مشاهده شده ایستگاه‌ها نشان داد که این مدل دارای توانایی برازش مناسب (R2=0.78-0.91) و با خطای پایین (RMSE=7.1-8.5 µg/m3) می‌باشد. رویه‌های بدست آمده نشان می‌دهند که تراکم غلظت در قسمت‌های مرکزی و شرقی شهر بالاتر است که این نشان دهنده تأثیر بادهای غربی بر جابجایی آلودگی در شهر می‌باشد. از طرفی با توجه به تراکم غلظت و ارتفاعات شهر تهران می‌توان دریافت که این ارتفاعات از خروج و جابجایی آلودگی جلوگیری نموده و باعث تراکم غلظت در مناطق شرقی شهر تهران می‌شوند. در رویه پائیزی نیز همانطور که می‌بینیم غلظت ماکزیمم در این روز، پهنه بیشتری از شهر را به نسبت دو روز دیگر در دو فصل قبلی در بر گرفته است که این به خاطر وجود دمای پائین در این فصل و اثر پدیده وارونگی بر تراکم غلظت در این فصول می‌باشد. فلذا موارد یاد شده نشان دهنده قدرت مدل پیشنهادی در اعمال جزئیات و پارامترهای مؤثر می‌باشد.

در ادامه این پژوهش، پیشنهاد می‌گردد مدل بیان شده در این پژوهش به‌صورت یک مدل پویا پیاده‌سازی شود و از آن در پایش آلودگی هوای شهر تهران استفاده گردد.

مراجع

[1] Soltani, A., Ahmadian, A., Esmaili, Y. (2010). " Using spatially weighted regression in assessment of spatial relations in a city" Armanshahr, 89 (4):pp. 99-110

[2] Rafipour, M. (2013). "Comparisons of efficiency of different neural networks in spatio-temporal prediction of air pollution of Tehran". MSc thesis, KNTU.

[3] Fathtabar, S., Alesheikh, A.A., Rangzan, K., Chinipardaz, R. (2011). "Zoning of air pollutants using of statistical models and GIS techniques, case study: Tehran", 5th conference of Environment engineering.

[4] Ghorbani, Kh. (2011). "GWR: a method for drawing of co-precipitations in Ghilan province", Water and Soil, 90 (3): pp. 752-743.

[5] Kavoosi, A., Sefidkar, R., Alavimajd, H., Imanzad, M., Noormoradi, H. (2013). "spatial analysis of air pollution in Tehran using autologistic regression, centralized autologistic regression, and pointer kriging , research journal of Ilam medicine university, 21 (7) : pp. 206-214

[6] Data and Research Unit (2013). "Annual report of air quality of Tehran in 2012". Air Quality Control Company, technical report, No. QM92/03/03/(U)/01.

[7] Burnham, K. P., Anderson, D. R.,(2004)." Multimodel Inference Understanding AIC and BIC in Model Selection". Sociological Methods & Research,.33(2): pp. 261-304.

[8] Brunsdon, C., Fotheringham, A.S. and Charlton, M.E.,(1996)."Geographically Weighted Regression: A Method for Exploring Spatial Nonstationarity". Geographical Analysis,. 28(4): pp. 281-298.

[9] Fotheringham A.S., Brunsdon C., and Charlton M.E. (2002). "Geographically weighted regression. Chichester" John Wiley & Sons.

[10] Gundogdu I., and ESEN, O.,(2010). "The importance of secondary variables for mapping of meteorological data." 3rd international conference on cartography and GIS. 15-20 June, Nessebar, Bulgaria.

[11] Kavousi, A., Sefidkar, R., Alavimajd, H., Rashidi, Y and Abolfazli Khonbi, Z. (2013). "Spatial analysis of CO and PM10 pollutants in Tehran city." Journal of Paramedical Sciences. VOL. 4. pp. 41-50.

[12] Propastin P., and Kappas M. (2008). "Reducing uncertainty in modeling the NDVI–precipitation relationship: a comparative study using global and local regression techniques." GISciRemote Sens. VOL. 45. pp. 47–67.

[13] Robinson, D.P., Lloyd, C.D., McKinley, J.M. (2013). "Increasing the accuracy of nitrogen dioxide (NO2) pollution mapping using geographically weighted regression (GWR) and geostatistics." International Journal of Applied Earth Observation and Geoinformation. VOL. 21. pp. 374–383.

[14] Tobler, W.R. (1970). A computer movie simulating urban growth in the Detroit region, Economic Geography VOL. 46. pp.234-240.

[15] Kumar Jha, D., Sabesan, M., Das, A., Vinithkumar, N.V., Kirubagaran, R.(2011). "Evaluation of Interpolation Technique for Air Quality Parameters in Port Blair, India." Universal Journal of Environmental Research and Technology. VOL. 1, NO. 3, pp. 301-310.

Archive of SID

www.SID.ir