

Tehran Air Pollution Estimation Improvement

Mahmoud Reza Delavar^{1*}, Amin Gholami², Gholam Reza Shiran³, Yousef Rashidi⁴, Gholam Reza Nakhaeizadeh⁵ and Kurt Fedra⁶, Smaeil Hatefi Afshar⁷

1 Center of Excellence in Geomatic Engineering, in Disaster Management, School of Surveying and Geospatial Engineering, College of Engineering, University of Tehran, Tehran, Iran.

2 Dept. of GIS, School of Surveying and Geospatial Engineering, College of Engineering, University of Tehran, Tehran, Iran.

3 Dept. of Transportation Eng., Faculty of Civil & Transportation Engineering, University of Isfahan, Isfahan, Iran.

4 Environmental Sciences Research Institute, Shahid Beheshti University, Tehran, Iran.

5 APL-Professor of Economics and Econometrics, Karlsruhe Institute of Technology, Institute of Economics Econometrics and Statistics, Karlsruhe, Germany.

6 Environmental Software & Services GmbH, Vienna, Austria.

7 Dept. of GIS, School of Surveying and Geospatial Engineering, College of Engineering, Tehran, Iran.

¹mdelavar@ut.ac.ir, ²amingholami328@gmail.com, ³gholam_shiran@yahoo.com, ⁴y_rashidi@sbu.ac.ir,

⁵gholamreza.nakhaeizadeh@kit.edu, ⁶kurt@ess.co.at and ⁷hafshar@ut.ac.ir

Corresponding author address: Mahmoud Reza Delavar, School of Surveying and Geospatial Engineering, College of Engineering, University of Tehran, Tehran, Iran, P.O.Box: 1439951154

Abstract- air pollution modeling is essential for environmental management. we attempted to select the optimum interpolation method for modeling air pollution in Tehran using some selected methods and proposed method. For this purpose, different interpolation methods such as inverse distances interpolation, polynomials, K nearest neighbors and various existing kriging methods were employed. Finally, by evaluating all the methods, the optimum interpolation method for air pollution modeling in Tehran was proposed and implemented. Among the above-mentioned methods, Kriging has had the best results. Then, using genetic optimization and particle swarm optimization methods, the clustering intervals of the Kriging method were optimized in both regular and irregular intervals to determine the optimal number of clusters and the distance between the clusters at regular intervals in the conventional methods. A mathematical variogram was used, in the irregular interval method. the same experimental variogram employed and the aim was to minimize the error of each of these methods. Finally, the error of the regular intervals with the optimal cluster numbers and spacing was less than those of all the mentioned methods. To test the proposed model, this model has been implemented for all months of the year (March 2016 – March 2017), with an average %64 accuracy improvement compared to that of the regular kriging method. The least error in the methods for interpolation of air pollution has been at 4 ($\mu\text{g} / \text{m}^3$), while an error of 1.8 ($\mu\text{g} / \text{m}^3$) in this study was achieved which resulted to the %55 accuracy improvement of the interpolation of air pollution, approximately. As a result, in this research, the best interpolation method has been presented in relation to previous studies for air pollution modeling . the proposed method can be applied as a complementary method for air pollution modeling where information scarcity exists in air pollution contamination.

Keywords- Interpolation, Air pollution modeling, PM_{2.5}, Kriging, Optimization.

بهبود برآورد میزان آلودگی هوای شهر تهران

محمودرضا دلاور^{۱*}، امین غلامی^۲، غلامرضا شیران^۳، یوسف رشیدی^۴، غلامرضا نخعی زاده^۵، کورت فدر^۶، اسماعیل هاتفی افشار^۷
*۱- استاد قطب علمی مهندسی نقشه برداری در مقابله با سوانح طبیعی، دانشکده مهندسی نقشه برداری و اطلاعات مکانی، پردیس دانشکده فنی، دانشگاه تهران، تهران، ایران

۲- گروه سیستم های اطلاعات مکانی، دانشکده مهندسی نقشه برداری و اطلاعات مکانی، پردیس دانشکده فنی، دانشگاه تهران، تهران، ایران
۳- استادیار گروه برنامه ریزی حمل و نقل، دانشکده فنی مهندسی، دانشگاه اصفهان، اصفهان، ایران

۴- استادیار گروه آلاینده های محیط زیست، پژوهشکده علوم محیطی، دانشگاه شهید بهشتی، تهران، ایران

۵- استاد دانشکده اقتصاد، دانشگاه کارلسروهه آلمان، کارلسروهه، آلمان

۶- شرکت خدمات و نرم افزار زیست محیطی GmbH، وین، اتریش

۷- گروه سیستم های اطلاعات مکانی، دانشکده مهندسی نقشه برداری و اطلاعات مکانی، پردیس دانشکده فنی، دانشگاه تهران، تهران، ایران

^{۱*} mdelavar@ut.ac.ir, ^۲ amingholami328@gmail.com, ^۳ gh.shiran@trn.ui.ac.ir, ^۴ y_rashidi@sbu.ac.ir, ^۵ nakhaezadeh@statistik.uni-karlsruhe.de, ^۶ kurt@ess.co.at and ^۷ hafshar@ut.ac.ir

* نشانی نویسنده مسئول: محمودرضا دلاور، تهران، دانشگاه تهران، دانشکده مهندسی نقشه برداری و اطلاعات مکانی، پردیس دانشکده فنی، کد پستی: ۱۴۳۹۹۵۱۱۵۴

چکیده- بسیاری از کلان شهرهای ایران به ویژه تهران با مشکل آلودگی هوا مواجه هستند که بخش عمده این آلودگی، ناشی از آلاینده PM_{2.5} می باشد لذا مدل سازی آلودگی هوا امری ضروری برای مدیریت محیط زیست می باشد. در این تحقیق از اطلاعات روزانه ماه های مختلف سال ۱۳۹۵ توسط حسگرهای سنسور آلودگی هوا مربوط به شرکت کنترل کیفیت شهرداری تهران و سازمان حفاظت محیط زیست برای مدل سازی آلودگی هوا استفاده شده است. ابتدا روش های مختلف درون یابی زمین آمار نظیر کریجینگ و غیر زمین آمار مانند فاصله معکوس وزندار، k نزدیکترین همسایه، چند جمله ای محلی و سراسری جت برآورد میزان آلاینده PM_{2.5} در نقاطی که مستقیماً میزان این آلودگی سنسور نشده استفاده شده است و نتایج با استفاده از روش ارزیابی متقاطع و شاخص خطای میانگین کمترین مربعات مقایسه شده اند. از بین روش های فوق کریجینگ بهترین نتیجه را در بر داشته است. سپس با استفاده از روش های بهینه سازی ژنتیک و ازدحام ذرات، فواصل خوشه بندی روش کریجینگ به دو صورت فواصل منظم و نامنظم بهینه سازی شده است که خطای فواصل منظم کمتر بوده است. سپس برای تست مدل پیشنهادی، این مدل برای تمامی ماه های مختلف سال پیاده سازی شده است که دقت در تمامی ماه ها به طور میانگین ۶۴٪ نسبت به روش کریجینگ معمولی بهبود یافته است. در نتیجه در این تحقیق بهترین روش درون یابی برای مدل سازی آلودگی هوا ارائه شده است. از این روش می توان به عنوان روشی مکمل در مواقعی که خلاء اطلاعات منابع آلودگی هوا وجود دارد برای تخمین میزان کیفیت هوا استفاده نمود.

واژه های کلیدی: درون یابی، مدل سازی میزان آلودگی هوا، PM_{2.5}، کریجینگ، بهینه سازی

۱- مقدمه

در دهه‌های اخیر، آلودگی هوا باعث افزایش نگرانی دولتها و ملت‌ها نسبت به عوارض (اثرات سو) بهداشتی آن شده است [۱]. بخش وسیعی از این آلودگی را ذرات معلق هوا که دارای اندازه‌ی بین ۰.۰۰۰۰۲ میکرون تا ۵۰۰ میکرون می‌باشند تشکیل می‌دهد [۲]. ذرات با قطر کمتر از ۲.۵ میکرون که $PM_{2.5}$ نامیده می‌شوند، می‌توانند به شش‌ها نفوذ کرده و اثرات مخربی برای سلامت فرد ایجاد نمایند [۳]. اکثر ذرات معلق موجود در هوا نسبت به آلاینده‌های دیگر گازی هوا، بیشتر موجب اختلالات تنفسی می‌شوند. از طرفی تاثیر متقابل ذرات معلق بر روی هم موجب تشدید اثرات سمی برخی از آلوده‌کننده‌ها خواهد شد (میری و همکاران، ۱۳۹۶). جدی‌ترین خطرات ممکن برای عملکرد ریوی و مشکلات قلبی و عروقی ناشی از ذرات $PM_{2.5}$ می‌باشد [۴]. هر ساله، حدود ۳.۵ میلیون مرگ و میر به علت بیماری‌های قلبی و عروقی و ۲۲۰۰۰۰ مرگ و میر نیز به دلیل سرطان ریه ناشی از ذرات $PM_{2.5}$ اتفاق می‌افتد [۵، ۶]. علم زمین‌آمار^۱ با فراهم کردن تخمین گرهای آماری مختلف نظیر کریجینگ، کمک می‌کند تا با استفاده از اطلاعات حاصله از نقاط نمونه برداری شده، به برآورد خصوصیت مورد نظر در نقاطی که نمونه برداری نشده‌اند پرداخته شود [۷]. در این رابطه، خصوصیات مکانی نقاط نزدیک به هم بسیار شبیه‌تر از نقاطی است که دور از هم قرار گرفته‌اند، بنابراین تصحیحات مکانی به صورت خودکار یکی از پیش‌نیازهای کاربرد زمین‌آمار می‌باشد. واریوگرام^۲ یا تحلیل تغییرنما، ابزار مفیدی برای تجزیه و تحلیل ساختار، مقیاس و شدت تغییرات مکانی متغیرهای ناحیه‌ای می‌باشد. واریوگرام، میانگین عدم شباهت بین مقادیر نمونه برداری نشده و نمونه برداری شده را اندازه‌گیری نموده و تصحیحات لازم را در فواصل مختلف به طور خودکار اعمال می‌کند. لذا مدل‌های زمین‌آماري مانند کریجینگ که در آنها با استفاده از واریوگرام‌ها درونیابی صورت می‌گیرد، کمک به سزایی در تخمین دقیق‌تر خصوصیت مورد نظر در نقاط نمونه برداری نشده ارائه می‌دهند [۸].

Levy و همکاران (۲۰۱۷) به کاهش خطای درونیابی آلودگی هوا پرداخته‌اند. به کمک تئوری بیزین و ارزیابی مقاطع، همبستگی بین خطای باقی مانده تخمین میزان غلظت آلاینده‌های NO_2 و NO_x مقادیر تخمین زده شده را به منظور بهینه سازی نقشه های غلظت آلاینده های هوا تعیین نموده و از این طریق درونیابی را بهبود بخشیدند. بر اساس نتایج این تحقیق پس از تصحیح مقدار خطای باقیمانده برای یک دوره دو ساله میزان آلودگی هوا از ۵.۱۲ (ppb^3) به ۴.۹۳ (ppb) کاهش پیدا کرده است [۹]. Li و Heap (۲۰۱۱) عملکرد ۶۱ روش درونیابی زمین‌آماري و غیر زمین‌آماري

را در ۵۱ مطالعه تطبیقی مقایسه کرده‌اند. تمرکز تحقیق فوق بر روی مطالعات زیست محیطی بوده است و آنها به این نتیجه رسیدند که به ترتیب روش کریجینگ معمولی (OK^4) و روش فاصله معکوس وزن دار (IDW^5) بیشترین استفاده را در این تحقیقات داشته‌اند و عملکردی بهتری نسبت به بقیه روش‌ها ارائه نموده‌اند. همچنین آنها در نتایج خود ذکر کرده‌اند که روش کریجینگ به جزء در چند مورد استثنا، عملکرد بهتری نسبت به روش‌های غیر زمین‌آماري داشته است [۱۰]. برای همین منظور در این تحقیق در میان روش‌های زمین‌آماري از روش کریجینگ برای مدلسازی آلودگی هوا استفاده شده است. Mercer و همکاران (۲۰۱۱) از دو روش رگرسیون کاربری زمین^۶ و کریجینگ سراسری^۷ برای پیش بینی غلظت NO_x در شهر لس‌آنجلس استفاده کرده‌اند. آنها در این تحقیق، علاوه بر پارامترهای هواشناسی و غلظت آلاینده‌ها، از برخی پارامترهای مکانی همچون فاصله از جاده‌ها، جمعیت، کاربری زمین و فاصله از مناطق ساحلی بهره برده‌اند. نتایج تحقیق فوق بیانگر دقت بهتر مدل کریجینگ سراسری نسبت به روش رگرسیون کاربری زمین در پیش‌بینی غلظت NO_x بوده است [۱۱].

حبیبی و آل شیخ (۱۳۹۵) به ارزیابی چهار روش درونیابی IDW ، EBK^8 ، GPI^9 و OK توسط دو شاخص اعتماد پذیری و زمان پرداختند. نتایج تحقیق مذکور نشان داد روش‌های OK و EBK با میانگین خطای ۱۴.۶ و ۱۵.۱ در شاخص اعتماد پذیری بهتر از دو روش دیگر عمل کردند و در شاخص زمان روش EBK عملکرد ضعیفی (۶ دقیقه و ۲۱ ثانیه)، اما دیگر روش‌ها از نظر زمانی در کسری از ثانیه و در حد مطلوبی عمل نموده‌اند. در نتیجه میتوان گفت روش OK ، هم از نظر اعتمادپذیری و هم از نظر زمان بهتر از بقیه الگوریتم‌ها عمل کرده است. این الگوریتم‌ها در یک محیط اینترنت اشیا برای داده‌های آلودگی هوا و جهت ارزیابی کیفیت سرویس‌های وب مکانی پیاده‌سازی و اجرا شده‌اند [۱۲].

میری و همکاران (۱۳۹۶) برای پهنه بندی آلودگی هوای ناشی از آلاینده‌ی $PM_{2.5}$ شهر مشهد، از روش‌های مختلف تحلیل فضایی نظیر OK ، IDW و UK^{10} استفاده کرده‌اند و به این نتیجه رسیدند که روش OK کمترین خطا و بیشترین ضریب همبستگی را نسبت به دو روش دیگر داشته است [۱۳]. بهاری و همکاران (۱۳۹۴) در تحقیق خود به پهنه بندی آلودگی ذرات معلق شهر تهران با استفاده از مدل‌های آماری در محیط سیستم‌های اطلاعات مکانی پرداخته‌اند آنها از روش رگرسیون وزندار جغرافیایی (GWR^{11}) برای مدلسازی استفاده کردند. آنها علاوه بر داده‌های آلودگی از پارامترهای هواشناسی (دما، سرعت و جهت باد)، کاربری زمین، راه‌های اصلی و آزادراه‌ها، جمعیت و ارتفاع استفاده کرده‌اند. آنها در تحقیق خود به

شوند که در اینصورت باز هم یک خطای درونیابی دیگر دخیل می‌شود. از طرف دیگر مطابق این تحقیقات، لازم است جهت برآورد آنی الودگی هوا ابتدا داده‌های کمکی جمع‌آوری یا محاسبه شود و سپس مقدار غلظت آلودگی هوا تخمین زده شود که این زمانبر است. در تحقیقات [۱۱]، [۱۲]، [۱۳]، [۱۴]، [۱۵] و [۱۸] به مقایسه روش‌های مختلفی پرداخته‌اند و روشی را بهبود نداده‌اند. در نتیجه این تحقیق در نظر دارد معایب تحقیقات مذکور را برطرف کند لذا در این تحقیق از هیچ داده کمکی دیگری استفاده نشده است و فقط از داده‌های آلودگی هوای آلاینده $PM_{2.5}$ استفاده شده است. از طرف دیگر این تحقیق در نظر دارد روش‌های مختلف را بررسی و مقایسه کند سپس بهترین روش را بهبود و توسعه بدهد.

در اکثر مطالعات انجام شده بهترین روش برای درونیابی آلودگی هوا، روش OK می‌باشد، لذا این تحقیق به دنبال آن است که این روش را بهبود دهد و با بقیه روش‌های زمین آمار و غیر زمین آمار مقایسه نماید. در این تحقیق بر طبق تحقیقات [۱۰]، [۱۳]، [۱۵] و [۱۸] فرض بر این است که از میان روش‌های درونیابی زمین آماری، روش کریجینگ و از میان روش‌های درونیابی غیر زمین آماری، روش IDW، چند جمله‌ای و نزدیکترین همسایه K-ام (KNN)، بهترین روش برای مدل سازی آلودگی هوا است. مقادیر غلظت آلاینده $PM_{2.5}$ توسط ایستگاه‌های شرکت کنترل کیفیت شهرداری تهران و سازمان حفاظت محیط زیست اندازه گیری شده اند. در نتیجه خروجی این تحقیق، ارائه یک مدل بهینه برای درونیابی و مدلسازی آلودگی هوای شهر تهران است که با توجه به این مدل می‌توان مقادیر آلاینده $PM_{2.5}$ در سطح شهر در نقاطی که میزان آلودگی هوا اندازه گیری نشده اند را تخمین زد.

این تحقیق به دنبال آن است که به سوالات زیر پاسخ دهد: کدام یک از روش‌های درونیابی زمین آماری و غیر زمین آماری برای مدلسازی آلودگی هوا، بهترین دقت را خواهد داشت؟ آیا می‌توان دقت واریوگرام تجربی را نسبت به واریوگرام‌های برازش داده شده ریاضی، بهبود بخشید؟

در تحقیق [۱۹] با استفاده از روش فازی ژنتیک جهت خوشه بندی فواصل داده‌ها جهت محاسبه واریوگرام‌ها استفاده کردند تا خطای تخمین را کاهش بدهند آنها با استفاده از روش فازی ژنتیک مشخص کردند که کدام فواصل با چه احتمالی در کدام خوشه باشد و در تحقیق [۲۰] با استفاده از الگوریتم بهینه‌سازی توده ذرات^{۲۴} پارامترهای محاسبه واریوگرام را بهبود داده‌اند، اما این تحقیق در نظر دارد تعداد خوشه‌ها و فواصل بینشان را تخمین بزند که نوآوری تحقیق در این مورد است.

در ادامه در بخش دوم به روش‌ها و مفاهیم مورد نیاز تحقیق پرداخته

خطای $7.1-8.5$ ($\mu g/m^3$) و ضریب همبستگی $0.75-0.80$ ($\mu g/m^3$) رسیده‌اند [۱۴].

Eslami و Ghaesmi (۲۰۱۸) برای تعیین بهترین روش درونیابی آلودگی هوای شهر تهران از ۷۱ روش مختلف که شامل مشتقات روش‌های نظیر IDW، GPI، LPI^{۲۲}، Kriging، RBF^{۱۳} و Cokriging می‌باشد پیاده‌سازی و مقایسه کرده‌اند. آنها در این تحقیق برای آلاینده $PM_{2.5}$ به بهترین خطای 7.17 ($\mu g/m^3$) با استفاده از روش کوکریجینگ رسیده‌اند. نتایج ۷۱ روش نشان داد که در اکثر موارد، روش زمین آماری بهتر از روش دقیق است [۱۵]. QI و همکاران (۲۰۱۸) یک روش یادگیری عمیق هوا^{۱۱} برای پیش‌بینی و درونیابی آلاینده‌های $PM_{2.5}$ ، PM_{10} ، CO، NO_2 و ازن شهر پکن توسعه داده‌اند که این روش نسبت به روش‌های یادگیری نظیر شبکه عصبی، رگرسیون لجستیک^{۱۵}، رگرسیون لاپلاسی^{۱۶}، شبکه عصبی مکرر^{۱۷} و $ARIMA^{18}$ خطای درونیابی و پیش‌بینی کمتری داشته است. در این تحقیق از داده‌های ساعتی دما، سرعت و جهت باد، فشار هوا و رطوبت هوا استفاده شده است [۱۶].

Nabavi و همکاران (۲۰۱۹) با استفاده از داده‌های تصاویر ماهواره‌ای ۱۰ کیلومتری هدف تاریک و عمق آبی (DB-DT^{۱۹}) عمق نوری آئروسول (AOD²⁰) و داده‌های ۱ کیلومتری اجرای چندزاویه‌ای تصحیح جوی (MAIAC^{۲۱}) به تخمین $PM_{2.5}$ شهر تهران پرداخته‌اند. آنها در این تحقیق علاوه بر داده‌های ذکر شده از داده‌های زاویه زینت خورشیدی، تراکم جاده، دید، ارتفاع، سرعت باد، رطوبت هوا، دما و روز از سال هم استفاده کردند و با استفاده از الگوریتم‌های یادگیری ماشینی^{۲۲} به خطای 17.52 ($\mu g/m^3$) و ضریب همبستگی 0.68 رسیده‌اند [۱۷].

Nunez-Alonso و همکاران (۲۰۱۹) به بررسی روش‌های زمین آمار و کریجینگ در شهر مادرید پرداختند. آنها در این تحقیق از سه بلوک منطقه‌ای با تراکم داده‌ی مختلف و از آنالیز مولفه اصلی^{۲۳} و خوشه‌بندی استفاده کردند که برای داده‌ها با تراکم‌های مختلف، هریک از روش‌ها دقت مناسبی دارند که به طور کلی روش‌های کریجینگ بهترین عملکرد داشته‌اند [۱۸].

از معایب پیشینه‌های مذکور می‌توان به طور خلاصه بیان کرد که در تحقیق [۹] نیازمند این است که اول درونیابی انجام شود سپس بعد از ایجاد خطای باقی مانده و محاسبه همبستگی بین این خطا و مقادیر تخمین زده شده، خطا را بهبود داد که زمانبر است از طرف دیگر خطای آن فقط ۳٪ بهبود پیدا کرده است. در تحقیقات [۱۱]، [۱۴]، [۱۵]، [۱۶]، [۱۷] و [۱۸] جهت تخمین الودگی هوا مطابق این روش‌ها، نیازمند این است که ابتدا داده‌های کمکی دیگر با خطای (دستگاهی، انسانی، محیطی و ..) اندازه‌گیری شوند و یا در صورت لزوم گاهی لازم است که این داده‌ها خودشان درونیابی

$$\hat{Z}(x_0) = \sum_{i=1}^n \lambda_i Z(x_i) + [1 - \sum_{i=1}^n \lambda_i] \mu \quad (2)$$

که در آن μ ، میانگین ثابت معلومات است. این پارامتر فرض می‌شود که در سرتاسر محدوده ثابت است و از میانگین داده‌ها محاسبه می‌شود [۲۴]. تعدادی از نقاط نمونه که توسط محدوده^{۲۷} تاثیر گذار در سمی واریوگرام^{۲۸} تعیین می‌شوند در این معادله برای تخمین استفاده می‌شوند [۲۴].

کریجینگ ساده یک شرط نا اریب^{۲۹} ندارد. در کریجینگ ساده لزوماً $1 - \sum_{i=1}^n \lambda_i$ برابر صفر نیست و هر چقدر این مقدار بزرگتر از صفر شود مقدار تخمین به میانگین نزدیکتر خواهد شد. به طور کلی افزایش نسبی این عبارت با ضعیف بودن منطقه نمونه‌برداری شده رابطه مستقیمی دارد [۲۶].

(ب) کریجینگ معمولی^{۳۰}

کریجینگ معمولی همانند کریجینگ ساده بوده با این تفاوت که در کریجینگ معمولی مقدار Z در معادله (۱) با جایگذاری میانگین کل μ با میانگین محلی پنجره همسایگی $\mu(x_0)$ تخمین زده می‌شود [۲۷، ۲۸].

$$[1 - \sum_{i=1}^n \lambda_i] = 0 \quad (3)$$

روش عمومی محاسبه کریجینگ بر اساس معادله (۲) [۲۷، ۲۸]:

$$\hat{Z}(x_0) = \sum_{i=1}^n \lambda_i Z(x_i) \quad (4)$$

در این معادله:

$\hat{Z}(x_0)$: مقدار محاسبه شده متغیر Z در نقطه x

λ_i : وزن یا اهمیت نسبت داده شده به متغیر Z در نقطه x_i

n : تعداد مشاهدات می‌باشد.

اولین گام در محاسبات کریجینگ، محاسبه واریوگرام است. واریوگرام تابعی است که بوسیله آن تغییرپذیری مکانی داده‌ها اندازه گیری می‌شود و در واقع مهمترین ابزار همبستگی فضایی داده‌ها است که اساس زمین آمار می‌باشد [۲۹]. با تقسیم واریوگرام بر دو، سمی واریوگرام بدست می‌آید که فهم آن آسانتر است. در روش‌های زمین آمار عملاً از این نوع واریوگرام استفاده می‌شود. واریوگرام که با علامت γ نشان داده می‌شود از طریق معادله (۵) محاسبه می‌شود [۲۹]:

$$\gamma(h) = \frac{1}{2n} \sum_{i=1}^n [Z_i(x) - Z_i(x+h)]^2 \quad (5)$$

که در آن مقادیر به صورت زیر می‌باشد:

شده است. در بخش سوم، روش تحقیق و روش پیشنهادی و نوآوری این تحقیق مورد بررسی قرار گرفته است و در بخش چهارم، منطقه مورد مطالعه، داده‌ها، نتایج پیاده‌سازی و یافته‌های تحقیق ارائه شده است. در نهایت در بخش پنجم، نتایج حاصله، تشریح شده و پیشنهاداتی نیز جهت تحقیقات آتی ارائه گردیده است.

۲- مواد و روش‌ها

تخمین زمین آماری، فرآیند تعیین مقدار یک کمیت در نقاط معلوم به کمک مقدار همان کمیت در سایر نقاط می‌باشد و در اوایل دهه ۱۹۱۰ در کشاورزی و در دهه ۱۹۳۰ در هواشناسی به کار برده شده است [۲۱]. اساس مدل‌های زمین آمار، تخمین کریجینگ در نقاط با مقادیر نامعلوم به عنوان یک عدد تصادفی با یک توزیع احتمالی مشخص در نقطه‌ای دلخواه از فضای مورد مطالعه است [۲۲]. قبل از استفاده از مدل‌های زمین آماری ابتدا داده‌ها از لحاظ نرمال بودن پراکنش بررسی می‌شوند و اگر داده‌ها به هر علتی از توزیع نرمال پیروی نکردند ابتدا به وسیله روشی مانند استفاده از لگاریتم، آنها را نرمال کرده و سپس در مدل زمین آماری مورد استفاده قرار می‌دهند. برای تبدیل این داده‌ها به یک نقشه با اطلاعات منطقه‌ای، از کریجینگ استفاده می‌شود [۲۳].

۲-۱- کریجینگ

کریجینگ روشی برای تخمین است که بر منطق میانگین متحرک وزن‌دار استوار می‌باشد و در مورد آن می‌توان گفت که بهترین تخمین گر خطی نا اریب^{۲۵} با کمترین پراش تخمین است. همه تخمین‌گرهای کریجینگ بر اساس معادله (۱) می‌باشند [۲۴]:

$$\hat{Z}(x_0) - \mu = \sum_{i=1}^n \lambda_i [Z(x_i) - \mu(x_0)] \quad (1)$$

که در آن پارامتر μ ، ثابت میانگین معلومات می‌باشد و فرض می‌شود که در کل دامنه ثابت است و توسط میانگین از داده‌ها محاسبه می‌شود [۲۴]. پارامتر λ_i وزن کریجینگ و پارامتر n تعداد نقاط نمونه استفاده شده در تخمین است که به اندازه پنجره همسایگی وابسته است و $\mu(x_0)$ میانگین نمونه‌های داخل پنجره همسایگی می‌باشد.

در این روش به هر یک از ایستگاه‌های درون و بیرون یک منطقه بر حسب فاصله و موقعیت آن وزن آماری تخصیص داده شده است تا پراش تخمین کمینه شود. روش‌های مختلف کریجینگ در ادامه تشریح شده‌اند [۲۴]:

الف) کریجینگ ساده^{۲۶}

معادله کریجینگ ساده در رابطه (۲) ارائه شده است [۲۴]:

می‌شود و از نسلی به نسل بعد تکامل می‌یابد. روند بهینه سازی از یک نسل که به صورت تصادفی تشکیل می‌شود شروع می‌شود. نسل اولیه شامل تعداد زیادی از راه حل‌های ممکن (کروموزم) است که معمولاً به شکل باینری برای مسئله بهینه سازی مورد نظر کدنویسی شده‌اند. در هر نسل تابع برازش^{۲۰} برای تمامی کروموزوم‌های آن نسل محاسبه می‌شود. این تابع برازش معمولاً تابع هدف مسئله بهینه سازی است. براساس مکانیزم نخبه گرایی، بهترین کروموزوم هر نسل به صورت مستقیم به نسل بعد منتقل می‌گردد. در طول فرآیند تولید مثل، کروموزوم‌های دارای مقدار تابع برازش کمتر به صورت اتفاقی در هر نسل بر اساس مکانیزم انتخاب، انتخاب می‌شوند و ژن‌های این کروموزوم‌ها با اعمال تغییراتی از سوی عملگرهای ژنتیک (ترکیب ضربی و جهش) به نسل بعد انتقال می‌یابند کروموزوم‌های تولید شده در حلقه‌های تکرار بعدی مورد استفاده قرار می‌گیرند. این فرآیند تا زمانی که به شرط توقف برسد یا به تعداد نسل‌های از قبل تعیین شده برسد اجرا می‌شود[۳۳].

۲-۳- الگوریتم بهینه سازی توده ذرات

فرآیند بهینه سازی با استفاده از الگوریتم PSO روشی است که به صورت گسترده در حل مسائل بهینه سازی مورد استفاده قرار می‌گیرد. این الگوریتم می‌تواند در یافتن روش‌های محاسباتی و عملکرد بهتر مورد استفاده قرار گیرد. الگوریتم PSO به سرعت تمام راه حل‌های ممکن هر مسئله بهینه سازی را محاسبه می‌نماید و بهترین جواب را به عنوان راه حل نهایی ارائه می‌دهد.

الگوریتم PSO یک الگوریتم جستجوی اجتماعی است که از روی رفتار اجتماعی و حرکات منظم جمعی پرندگان و ماهی‌ها اقتباس شده است. علیرغم توانایی محدود هر ذره در یافتن بهترین الگو، رفتار جمعی آنها قابلیت و توانایی زیادی در پیدا کردن بهترین مسیر یا به عبارت دیگر بهترین جواب در مسائل بهینه سازی را دارد زیرا تغییر موقعیت هر ذره بر اساس تجربه شخصی ذره در حرکات قبلی و تجربه ذرات همسایه صورت می‌گیرد. در واقع هر ذره از برتری خود نسبت به ذرات همسایه و همچنین نسبت به کل ذرات آگاه است[۳۴].

۳- روش تحقیق

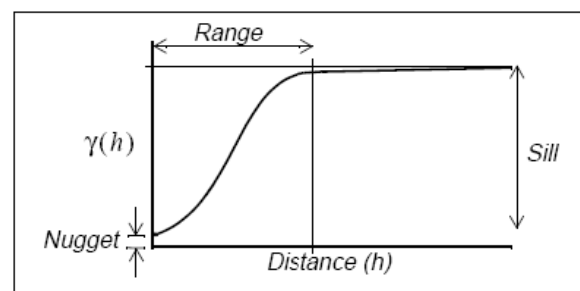
مطابق شکل (۲) به منظور مدل سازی آلودگی هوای شهر تهران ابتدا داده‌های حسگرهای آلودگی هوا جمع آوری شد. بعد از نرمال سازی داده‌ها با استفاده از تبدیلات لگاریتمی، به بررسی روند تغییرات پارامترهای مورد مطالعه پرداخته شد.

n: تعداد جفت نمونه‌ها به ازاء هر فاصله h

$Z_i(x)$: مقدار متغیر در نقطه x

$Z_i(x+h)$: مقدار متغیر در نقطه‌ای به فاصله h از نقطه x

با افزایش مقدار h، مقدار واریوگرام تا فاصله معینی اضافه می‌شود و پس از آن به حد ثابتی می‌رسد که حد آستانه^{۳۱} گفته می‌شود. به فاصله بین نمونه‌ها که از آن به بعد مقادیر متغیر در نقاط مجاور تاثیر چندانی بر یکدیگر ندارند و با افزایش بیشتر فاصله، مقدار واریوگرام، تفاوت معنی داری نمی‌کند، دامنه یا شعاع تاثیر^{۳۲} گفته می‌شود. به مقدار واریوگرام به ازاء $h=0$ اثر قطعه ای یا واریانس تصادفی^{۳۳} گفته می‌شود[۲۷]. عواملی مانند خطای نمونه برداری یا تغییرات در مقیاس کوچک می‌تواند باعث عدم شباهت و همبستگی بین مقادیر نمونه برداری شده‌ای که در فاصله بسیار نزدیک به هم هستند، گردد و این عوامل باعث عدم پیوستگی واریوگرام در نقطه شروع می‌شود. در شکل (۱) یک مدل واریوگرام ریاضی نمایش داده شده است[۳۰].

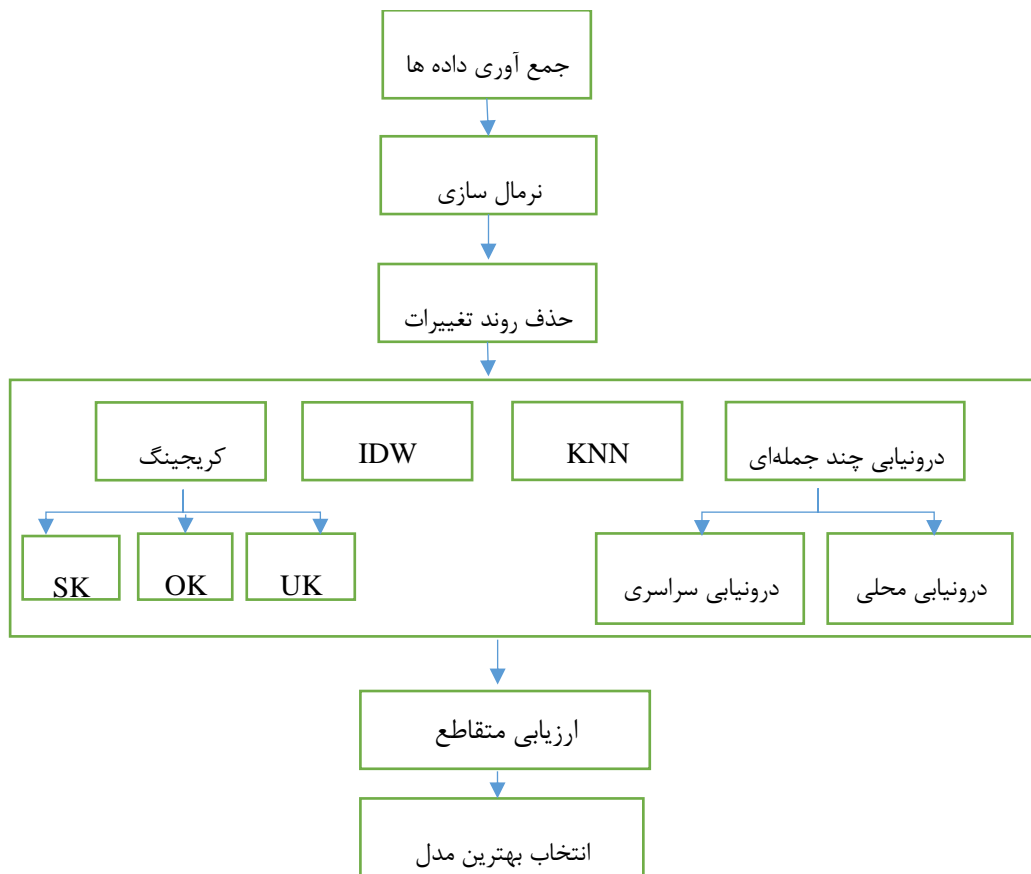


شکل ۱: یک مدل واریوگرام ریاضی با اثر قطعه‌ای، آستانه و شعاع تاثیر[۲۴].

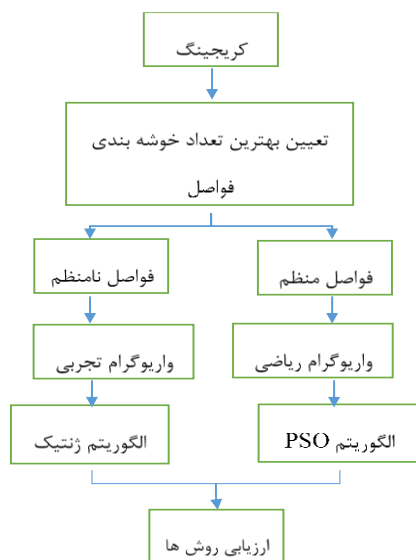
مدل واریوگرام و تخمین، برای تجزیه و تحلیل ساختاری و درونیابی مکانی بسیار مهم است[۲۵]. مدل‌های واریوگرام ممکن است مدل‌هایی ساده شامل قطعه‌ای^{۳۴}، نمایی^{۳۵}، کروی^{۳۶}، گوسین^{۳۷}، خطی^{۳۸} و مدل توانی^{۳۹} یا مجموع تو در تو از یک یا چند مدل ساده باشد[۲۰، ۲۵، ۳۱].

۲-۲- الگوریتم ژنتیک

الگوریتم ژنتیک توسط Holland (۱۹۹۲) توسعه داده شده است. الگوریتم ژنتیک یک روش جستجوی انطباق پذیر است می‌تواند به طور مناسب بدون گیر افتادن در بهینه‌های محلی، تمام فضای جستجو را بررسی نماید. این الگوریتم بر اساس اصول فرآیند تکامل ژنتیکی است که در آن از تئوری تکامل طبیعی الهام گرفته شده است[۳۲]. الگوریتم ژنتیک سنتی یا ساده از یک نسل اولیه شروع



شکل ۲: روش تحقیق جهت انتخاب بهترین مدل درونیابی آلودگی هوا



شکل ۳: روش پیشنهادی تحقیق جهت مدلسازی میزان آلودگی هوا.

در روش کریجینگ تجربی اگر n مشاهده داشته باشیم به تعداد $\frac{n(n-1)}{2}$ فاصله h_{ij} بین مشاهده i و j خواهیم داشت. هریک از این h_{ij} ها باید در یکی از این خوشه‌ها قرار گیرد. با مجموعه ای از این

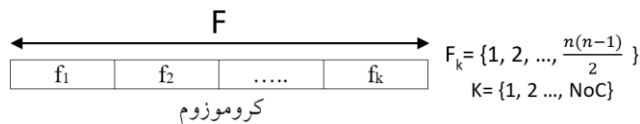
قبل از اجرای روش‌های درونیابی برای اینکه نقشه‌های تخمین مناسبتری تولید گردد، روند داده‌ها حذف شد. در این تحقیق سعی شد که بهترین روش درونیابی برای مدلسازی آلودگی هوای شهر تهران بر مبنای بهینه‌سازی روش‌های موجود انتخاب و استفاده گردد. برای این منظور روش‌های مختلف درونیابی نظیر درونیابی معکوس فاصله، چند جمله‌ای‌ها، K نزدیک‌ترین همسایه و انواع روش‌های کریجینگ موجود (ساده، معمولی، سراسری و...) و همچنین روش‌های کریجینگ پیشنهادی که بهینه‌سازی شده روش‌های کریجینگ است به کار گرفته شد. در نهایت با ارزیابی تمامی روش‌ها، بهترین مدل درونیابی جهت مدلسازی آلودگی هوای شهر تهران پیشنهاد و پیاده سازی شد. سپس برای ارزیابی، روش پیشنهادی برای مجموعه داده‌های مختلف ماه‌های سال ۱۳۹۵ برای آلاینده $PM_{2.5}$ استفاده شده است. روش تحقیق جهت بهبود روش کریجینگ بر اساس شکل (۲) انجام شده است.

در شکل (۳) مراحل پیشنهادی ذکر شده است، بدین صورت که از بین روش‌های موجود کریجینگ، بهترین مدل انتخاب و بهینه‌سازی گردید. از آنجا که در کریجینگ نیاز به خوشه‌بندی فواصل می‌باشد تعداد یا فاصله این خوشه‌بندی‌ها می‌تواند مهم باشد.

که در آن n ، تعداد نقاط نمونه، $Z_{i,act}$ ارزش نقطه معلوم i و $Z_{i,est}$ برآورد نقطه معلوم i است. بهترین برآورد باید کمترین RMSE را داشته باشد.

مطابق شکل (۳) برای فواصل نامنظم از الگوریتم ژنتیک جهت بهبود کریجینگ استفاده شده است. در این بخش ابتدا مطابق روش ارزیابی متقاطع یکی از داده‌ها از داده‌های چک حذف می‌شود سپس با استفاده از کریجینگ مقدار آلودگی هوای این داده بر اساس داده‌های چک مطابق رابطه (۴) محاسبه می‌شود این روند برای تمامی داده‌ها بکار برده می‌شود سپس با استفاده از رابطه (۶) مقدار خطا محاسبه می‌شود، سپس با استفاده از الگوریتم ژنتیک هدفمان این است که برای هر NoC یک ترکیب F_k مشخص شود به شرط آنکه کمترین خطا در کریجینگ حاصل شود.

کروموزم‌های این تحقیق همان ترکیبات مختلف F و تعداد ژن هر کروموزم برابر با NoC می‌باشد که در شکل (۴) نشان داده شده است و جمعیت اولیه کروموزم‌ها تصادفی انتخاب شده است در جدول (۱) پارامترهای اساسی الگوریتم ژنتیک مورد استفاده در این تحقیق نمایش داده شده است.



شکل ۴: ساختار کروموزوم مورد استفاده این تحقیق در الگوریتم ژنتیک

جدول ۱: پارامترهای اساسی الگوریتم ژنتیک مورد استفاده در تحقیق

پارامتر	مقدار
Population Size	۱۵
Number of Generations	۱۰۰
Crossover rate	۰.۸
Mutation rate	۰.۰۱
Crossover Method	One point
Elite count	۱

تابع برازش این الگوریتم در رابطه (۷) آمده است. برای هر کروموزم در نسل‌های متفاوت، تابع برازش مشخص می‌شود. براساس مکانیسم نخبه گرایی، بهترین کروموزم هر نسل به صورت مستقیم به نسل بعد منتقل می‌گردد. در طول فرآیند تولید مثل، کروموزوم‌های دارای مقدار تابع برازش کمتر به صورت اتفاقی در هر نسل بر اساس مکانیسم انتخاب، انتخاب می‌شوند و ژن‌های این کروموزوم‌ها با اعمال تغییراتی از سوی عملگرهای ژنتیک (ترکیب ضربی و جهش) به نسل بعد انتقال می‌یابند کروموزوم‌های تولید شده در حلقه‌های تکرار بعدی مورد استفاده قرار می‌گیرند. این فرآیند تا زمانی که به شرط توقف برسد یا به تعداد نسل‌های از قبل تعیین شده برسد اجرا می‌شود.

$$\text{Fitness Function} = \min\{\text{RMSE}\} \quad (7)$$

h_{ij} ها در هریک از این خوشه‌ها، یک واریوگرام تجربی طبق رابطه (۵) برای هر خوشه بدست می‌آید. می‌توان گفت دقت درونیابی به تعداد این خوشه‌ها و همچنین فاصله این خوشه‌ها یا به عبارت دیگر تعداد فواصل h_{ij} در هر خوشه وابسته می‌باشد. تعداد خوشه‌ها را با نماد NoC^{41} و تعداد مشاهدات h_{ij} که در خوشه k قرار می‌گیرند با نماد F_k نمایش داده شده است. در این صورت $k=1, \dots, NoC$ خواهد بود که در تحقیق حاضر پس از چندبار اجرای برنامه، NoC برابر با ۵، ۸، ۱۰، ۱۲ یا ۱۵ در نظر گرفته شد و فاصله خوشه‌ها به دو دسته منظم و نامنظم تقسیم بندی گردید. به طور معمول در روش کریجینگ فواصل به طور منظم تقسیم می‌شود و برای هر یک از این فواصل منظم، یک واریوگرام تجربی طبق رابطه (۵) بدست می‌آید. طبق شکل (۱) نمودار واریوگرام باید صعودی باشد تا اینکه از یک نقطه به بعد به یک حالت ثابت برسد. حال برای رسیدن به این هدف از مدل‌های واریوگرام ریاضی نظیر مدل کروی، گوسی، نمایی و دایره‌ای استفاده گردید. در این روش نیاز به دانستن پارامترهای مربوطه به هریک از این مدل‌ها می‌باشد که با تغییر مقادیر این پارامترها، دقت این روش تغییر خواهد کرد. بدین منظور در این تحقیق مقادیر پارامترها توسط الگوریتم PSO به گونه ای مشخص گردید که بهترین دقت برای درونیابی حاصل شود و همچنین در این روش تعداد خوشه‌ها متغیر در نظر گرفته شد.

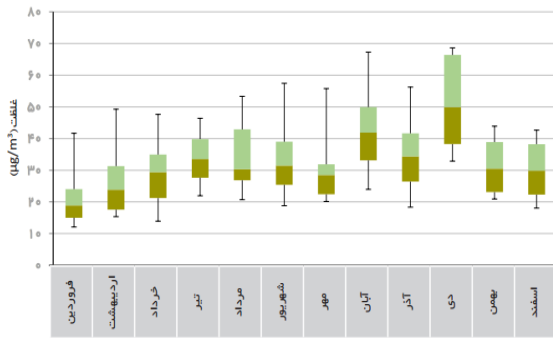
در حالت دوم درونیابی کریجینگ این تحقیق، فواصل به طور نامنظم در دامنه تغییرات آن تقسیم بندی گردید. در این روش تعداد خوشه‌ها متغیر و فواصل خوشه‌ها نامشخص و نامنظم است. از آنجائیکه با تغییر فاصله خوشه K ، تابع تناسب (F_k) نیز تغییر خواهد کرد، مقادیر F_k نیز در این روش متغیر می‌باشد. در این تحقیق مقادیر F_k را با استفاده از الگوریتم ژنتیک بهینه‌سازی نموده و به دنباله‌ی آن بهترین تعداد خوشه‌ها (NoC) و بهترین فاصله برای هر خوشه تعیین می‌شود تا دقت مورد نظر حاصل شود.

برای ارزیابی و کنترل قابلیت و اعتبار سنجی پارامترهای مورد استفاده در تخمین از روشی موسوم به ارزیابی متقاطع^{۴۲} استفاده شده است. این روش شامل حذف متوالی نمونه‌ها و برآورد مجدد آنها به روش کریجینگ و با استفاده از سایر نمونه‌ها و مدل برازش شده بر اساس واریوگرام تجربی است. سپس از تفاضل مقادیر واقعی و برآورد شده برای ارزیابی برآوردها استفاده گردید. در این تحقیق از معیار خطای جذر میانگین مربعات^{۴۳} استفاده شد که از رابطه (۶) محاسبه گردید [۲۹]:

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (Z_{i,act} - Z_{i,est})^2} \quad (6)$$

۴- پیاده‌سازی و کار عملی

در این تحقیق شهر تهران به عنوان منطقه مورد مطالعه در نظر گرفته است. در شکل (۵) نقشه مناطق شهر تهران به همراه ایستگاه‌های سازمان حفاظت محیط زیست و شرکت کنترل کیفیت شهرداری تهران نشان داده شده است. داده‌های آلودگی هوای شهر تهران مربوط به ایستگاه‌های شرکت کنترل کیفیت شهر تهران و سازمان حفاظت محیط زیست برای آلاینده PM_{2.5} جمع آوری شده است.



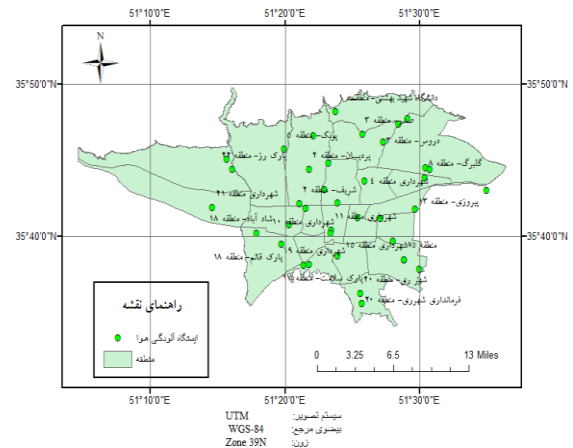
شکل ۶: تغییرات ماهیانه غلظت آلاینده PM_{2.5} در سال ۱۳۹۵ در شهر تهران

همانطور که در شکل (۶) مشخص شده در ماه فروردین مقارن با تعطیلات نوروزی با کاهش ترافیک و شرایط مناسب جوی، کمترین غلظت ماهیانه این آلاینده اتفاق افتاده است و میانگین ماهیانه غلظت آلاینده با شروع فصل گرم و بروز پدیده گرد و غبار در شهر، افزایش یافته است. البته بیشترین مقادیر در غلظت ماهیانه به ترتیب در ماه‌های دی و آبان دیده می‌شود و علت آن برودت هوا، افزایش میزان پایداری جوی و وارونگی دما است که به انباشت آلاینده‌ها در هوای شهر منجر شده است.

آن دسته از داده‌هایی که ضریب چولگی بالای ۰.۵ داشته باشند، غیر نرمال تشخیص داده شده و برای نرمال سازی این داده‌ها از روش لگاریتم‌گیری استفاده شده است. در غیر اینصورت اگر این ضریب کمتر از ۰.۵ باشد، داده‌ها احتیاجی به نرمال سازی نخواهند داشت. پس از حذف روند داده‌ها، از مدل‌های مختلف درونیابی برای مدلسازی آلودگی هوا استفاده شد.

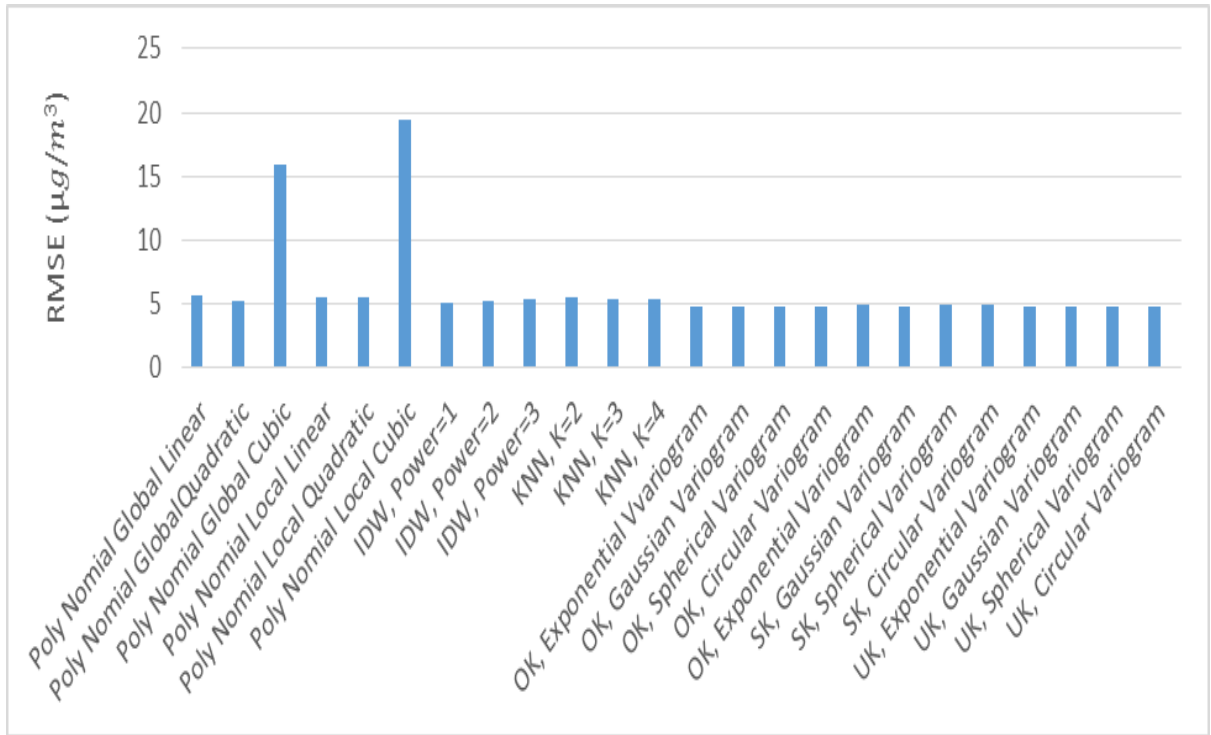
برای تعیین بهترین مدل درونیابی آلودگی هوا، از داده روزانه آلاینده PM_{2.5} در فروردین ماه استفاده شده است. پس از آماده سازی داده-ها، در این تحقیق برای مدلسازی آلودگی هوای شهر تهران، روش-های درونیابی زمین‌آماري نظیر کریجینگ و غیر زمین‌آماري نظیر KNN, IDW و چندجمله‌ای‌ها در محیط GIS تحلیل شده‌اند که نتایج آن در شکل (۷) نشان داده شده است.

با توجه به نتایج شکل (۷)، بهترین نتیجه روش چندجمله‌ای، مربوط به چندجمله‌ای سراسری درجه دوم و بهترین نتیجه روش IDW، مربوط به توان یک و بهترین نتیجه روش KNN، مربوط به K=4 و بهترین نتیجه روش کریجینگ مربوط به کریجینگ معمولی با مدل واریوگرام ریاضی دایروی می‌باشد که نقشه درونیابی هر یک از این روش‌ها در شکل (۸) ارائه شده است.

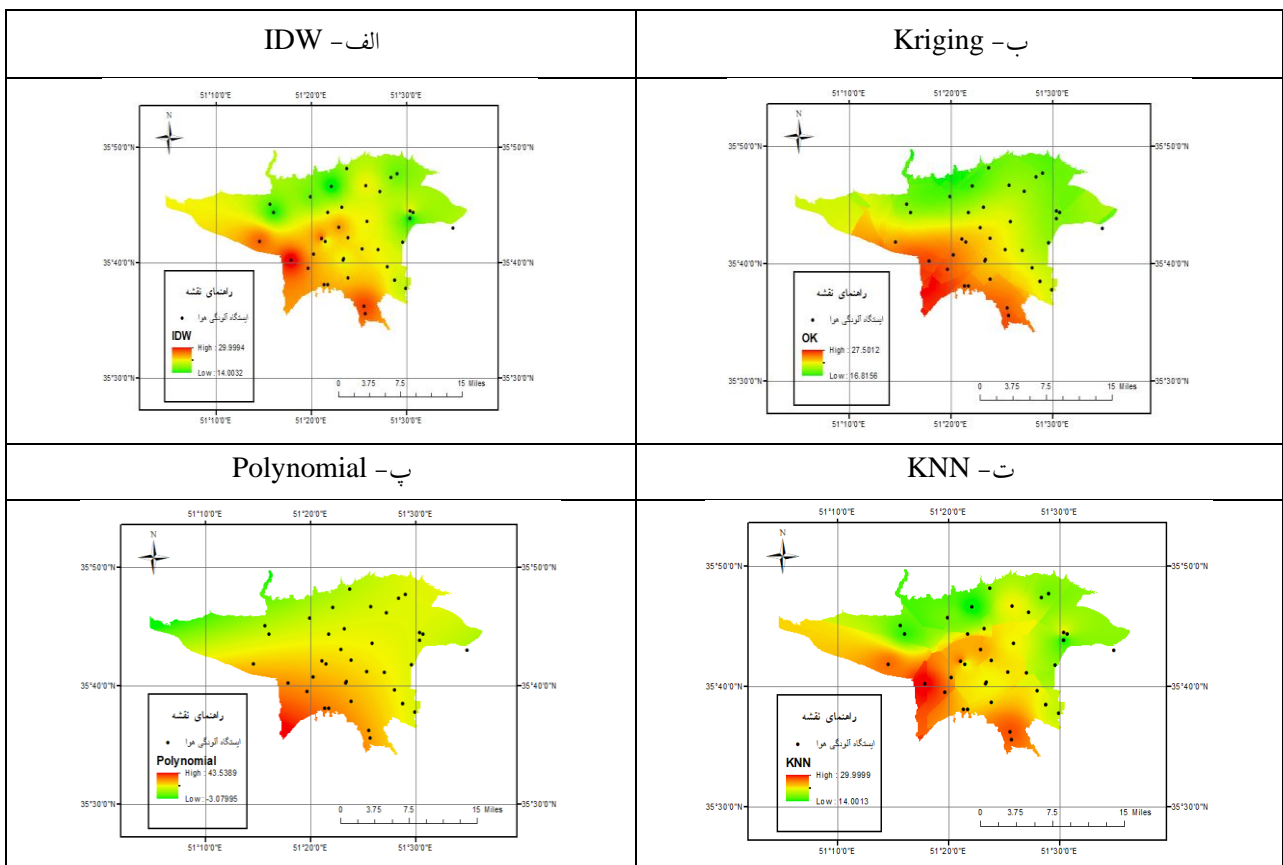


شکل ۵: مناطق شهر تهران و ایستگاه‌های سنجش آلودگی هوا.

برای مدلسازی آلودگی هوا برای تمامی ماه‌ها، از داده‌های روزانه سال ۱۳۹۵ استفاده شده است. در شکل (۶) غلظت میانگین آلاینده PM_{2.5} در ماه‌های مختلف سال ۱۳۹۵ نشان داده شده است. این آلاینده که نماینده طیف وسیعی از ذرات بسیار ریز با ابعاد متفاوت با قطر کمتر از ۲.۵ میکرون می‌باشد، بیشتر از فرآیندهای احتراقی به‌ویژه در وسایل نقلیه موتوری نظیر خودروهای دیزل تولید می‌شود و در صورت حاکمیت شرایط جوی پایدار برای چند روز متوالی، در هوای شهر انباشت می‌شود. شایان ذکر است، این آلاینده در اکثر قریب به اتفاق روزها، آلاینده شاخص به شمار می‌رود. آلودگی هوا در ماه‌های مختلف سال رفتارهای مختلفی از خود نشان می‌دهد که در شکل (۶) تغییرات غلظت آلاینده PM_{2.5} آمده است.



شکل ۷: نتایج روش‌های درونیابی برای مدلسازی آلودگی هوا شهر تهران



شکل ۸: نقشه میزان آلاینده PM_{2.5} در شهر تهران با استفاده از روش‌های الف - IDW، ب - Kriging، پ - Polynomial، ت - KNN

جدول ۵- مدل واریوگرام کروی به ازای NOC های مختلف توسط الگوریتم PSO

NoC	Nugget effect (m)	Sill (m)	Range (m)	RMSE ($\mu\text{g}/\text{m}^3$)
5	0.0001	0.505	8280.5	3.65
10	0.870	0.727	19736.6	5.25
12	0.001	0.654	16457.2	4.32
15	0.705	0.909	19971.7	5.06

طبق نتایج بدست آمده از مدل های واریوگرام ریاضی، کمترین خطا برای NoC=10 برای هر دو مدل واریوگرام گوسی و نمایی به دست آمده است و همچنین برای داده های آلودگی هوا بهترین مدل برای تمامی مقادیر NoC، مدل واریوگرام گوسی می باشد.

در این تحقیق یک مدل پیشنهادی ارائه شده است که در آن برای فواصل نامنظم خوشه ها، بهترین ترکیب F به ازای NOC های مختلف برای مدل واریوگرام بدست آمده است که به ازای هر یک از NOC ها، یک الگوریتم ژنتیک پیاده سازی شده است تا کمترین RMSE به ازای ترکیب F بدست آید.

در نهایت بهترین کروموزم برای هر یک از NOC ها بعد از ۱۰۰ نسل که کمترین تابع برازش را داشته باشد انتخاب شده و در جدول (۶) ارائه شده است.

جدول ۱- مدل پیشنهادی به ازای NOC های مختلف توسط الگوریتم ژنتیک

NoC	Best F	RMSE ($\mu\text{g}/\text{m}^3$)
5	{1,53,35,7,9}	1.815
8	{1,16,21,19,20,10,11,7}	4.088
10	{1,10,12,15,16,19,9,7,11,5}	2.381
12	{1,5,10,14,8,15,17,9,5,8,8,5}	2.625
15	{2,4,5,8,10,8,12,11,13,6,4,6,7,4,5}	3.295

در جدول (۶) همانطور که مشاهده می گردد به ازای داده های مختلف در روزهای مختلف، این ترکیب متفاوت خواهد بود. لذا جهت مدل سازی آلودگی هوا، مدل پیشنهادی بهترین نتیجه را در بر داشته است. بدین منظور در این تحقیق جهت درون یابی آلاینده $\text{PM}_{2.5}$ در ماه های مختلف سال ۱۳۹۵ از روش پیشنهادی استفاده شده است که نتایج آن در شکل (۹) نمایش داده شده است.

همانطور که در شکل (۹) مشاهده می شود برای تمامی ماه های سال حتی در ماه های فصل زمستان که بیشترین سهم آلودگی هوای سال در این فصل رخ می دهد، مدل پیشنهادی این تحقیق کمترین خطا را در برداشته است. به طور میانگین دقت روش کریجینگ معمولی در مدل پیشنهادی برای مدل سازی آلودگی هوا ۶۴٪ بهبود یافته است.

با توجه به نتایج شکل (۷) و (۸)، برای مدل سازی آلودگی هوای شهر تهران، روش کریجینگ معمولی دارای بهترین برازش است. در نتیجه در این تحقیق سعی شده است با استفاده از روش های پیشنهادی مذکور، دقت روش کریجینگ معمولی بهینه شود. برای روش فواصل منظم، مدل های واریوگرام ریاضی نظیر مدل کروی، گوسی، دایره ای و نمایی برای تعداد خوشه های مختلف توسط الگوریتم PSO، بهترین مقادیر حد آستانه، دامنه و اثر قطعه ای در نظر گرفته شده است. در الگوریتم PSO استفاده شده این پژوهش هر ذره شامل سه بعد حد آستانه، دامنه و اثر قطعه ای می باشد که با بدست آوردن مقادیر تقریبی این سه پارامتر و بیشینه و کمینه این ابعاد، توسط الگوریتم PSO تخمین زده شده است. در نتیجه در این الگوریتم بهترین تابع برازش برای هر ذره با سه بعد، تخمین زده شده است.

تابع برازش پیشنهادی الگوریتم PSO در این تحقیق، مطابق رابطه (۷) است. که برای هر ذره بر اساس سه بعد (سه پارامتر) تابع برازش بدست می آید و در نهایت کمترین میزان RMSE یا همان بهترین تابع برازش ذرات، برای مدل های مذکور در حالت های مختلف NoC توسط الگوریتم PSO بدست آمده است که در جدول (۲) تا (۵) ارائه شده است.

جدول ۲: مدل واریوگرام نمایی به ازای NOC های مختلف توسط

الگوریتم PSO

NoC	Nugget effect (m)	Sill (m)	Range (m)	RMSE ($\mu\text{g}/\text{m}^3$)
5	0.278	0.617	15562.1	3.09
10	0.072	0.789	10979.2	2.37
12	0.050	0.476	19139.8	4.21
15	0.252	0.672	15400.81	3.99

جدول ۳: مدل واریوگرام گوسی به ازای NOC های مختلف توسط

الگوریتم PSO

NoC	Nugget effect (m)	Sill (m)	Range (m)	RMSE ($\mu\text{g}/\text{m}^3$)
5	0.262	0.288	16418.05	3.79
10	0.445	0.702	15334.4	2.38
12	0.225	0.940	17551.3	3.33
15	0.516	0.766	15345.7	2.83

جدول ۴: مدل واریوگرام دایروی به ازای NOC های مختلف توسط

الگوریتم PSO

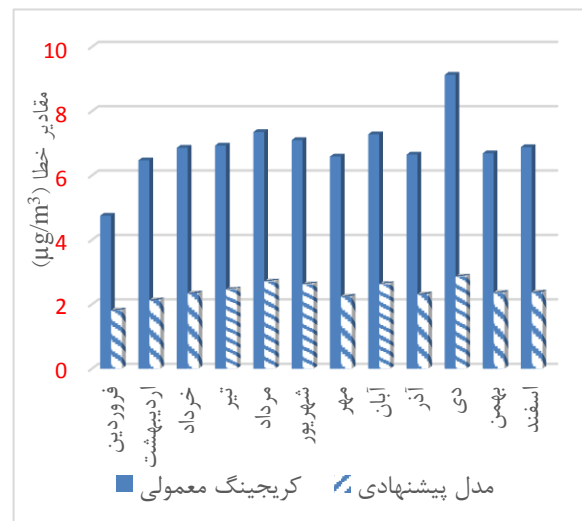
NoC	Nugget effect (m)	Sill (m)	Range (m)	RMSE ($\mu\text{g}/\text{m}^3$)
5	0.766	0.01	11479.3	5.48
10	0.646	0.01	10897.9	5.48
12	0.014	0.475	15197.8	10.09
15	0.792	0.01	12182.6	5.48

۵- نتیجه‌گیری و پیشنهادات

در این تحقیق در جهت مدل‌سازی آلودگی هوای شهر تهران با لحاظ نمودن آلاینده $PM_{2.5}$ از بین روش‌های مختلف درونیابی آلودگی هوا، روش کریجینگ معمولی که از جمله روشهای زمین آمار است، دقت بهتری را نسبت به سایر روشهای درونیابی ارزیابی شده ارائه نموده است.

در این تحقیق سعی شده است با استفاده از روش‌های پیشنهادی مذکور، دقت روش کریجینگ معمولی بهینه شود. برای روش فواصل منظم، مدل‌های واریوگرام ریاضی نظیر مدل کروی، گوسی، دایره‌ای و نمایی برای NoC های مختلف توسط الگوریتم PSO، بهترین مقادیر آستانه، دامنه یا شعاع تاثیر و اثر قطعه ای یا واریانس تصادفی در نظر گرفته شده است و در نهایت معیار RMSE برای مدل‌های مذکور در حالت‌های مختلف NoC بدست آمده است. این روش برای داده‌های ماه‌های مختلف سال ۱۳۹۵ مورد تست قرار گرفته است و برای هر ماه ترکیب F مجزا ایجاد شده است که دقت درونیابی کریجینگ تجربی توسط این مجموعه F_k ها بهبود پیدا کرده است. به طور میانگین برای تمامی ماه‌های سال، دقت مدل پیشنهادی به نسبت روش کریجینگ معمولی ۶۴٪ بهبود یافته است در صورتی که در تحقیق [۹] فقط ۳٪ خطای درونیابی بهبود پیدا کرده است. در نتیجه می‌توان ادعا نمود که واریوگرام تجربی می‌تواند نسبت به واریوگرام‌های ریاضی با استفاده از بهینه‌سازی فواصل و تعداد خوشه‌ها، دقت بهتری را بدست آورد.

نتایج این تحقیق با نتایج تحقیق [۹]، [۱۰] و [۱۳] همخوانی دارد بدین صورت که مدل OK بهترین دقت را نسبت به سایر روش‌ها به همراه داشته است. در تحقیق [۱۴] که به مدل‌سازی آلودگی هوا با استفاده از داده‌های پارامترهای هواشناسی (دما، سرعت و جهت باد)، کاربریهای ۹ گانه، راههای اصلی و آزادراه‌ها، جمعیت و ارتفاع پرداخته‌اند، خطای $7.1-8.5 (\mu g/m^3)$ حاصل شده است در صورتی که در این تحقیق فقط با استفاده از داده‌های آلاینده $PM_{2.5}$ خطای $1.81-2.76 (\mu g/m^3)$ حاصل شده است. کمترین خطا در روش‌های درونیابی آلودگی هوا در این زمینه ۴ $(\mu g/m^3)$ بوده است که در این تحقیق، خطای $۱.۸ (\mu g/m^3)$ حاصل شده است که در نتیجه دقت درونیابی آلودگی هوا حدوداً ۵۵٪ بهبود یافته است. در نتیجه می‌توان ادعا نمود که مدل پیشنهادی این تحقیق، مدل مناسبی برای مدل‌سازی آلودگی هوا است. از این روش می‌توان به عنوان روشی مکمل در مواقعی که خلاء اطلاعات منابع آلودگی هوا وجود دارد برای تخمین میزان کیفیت هوا استفاده نمود. برای تحقیقات آتی، پیشنهادات زیر جهت بهبود دقت محاسبه میزان آلودگی هوا ارائه می‌گردند:



شکل ۹: مقایسه خطای مدل کریجینگ معمولی و مدل پیشنهادی برای مدل‌سازی آلودگی هوا

جهت ارزیابی و تست این مدل پیشنهادی برای داده‌های شهری دیگر از داده‌های ماه می ۴۴ سال ۲۰۱۶ شهر هیوستون^{۴۵} در ایالت تگزاس آمریکا، استفاده شده است. این شهر از ۹ ایستگاه سنجش میزان آلودگی هوای آلاینده $PM_{2.5}$ برخوردار است که از تمامی این ایستگاه‌ها جهت ارزیابی مدل پیشنهادی استفاده شده است. به طور میانگین خطای روش کریجینگ معمولی و مدل پیشنهادی برای تمامی ایستگاه‌های هیوستون به ترتیب معادل ۶.۸۴ و ۳.۲۸ بدست آمده است که در نتیجه مدل پیشنهادی عملکرد بهتری برای تخمین میزان آلودگی هوای شهر هیوستون داشته و خطا را ۵۲٪ برای این شهر بهبود داده است. در نتیجه می‌توان ادعا نمود که این مدل پیشنهادی می‌تواند برای شهرهای دیگر هم استفاده شود.

در جدول (۷) روش پیشنهادی این تحقیق برای $NoC=5$ با روش‌های پیشنهادی تحقیقات [۱۹] و [۲۰] برای داده‌های این تحقیق پیاده‌سازی و مقایسه شده است.

جدول ۷- مقایسه مدل پیشنهادی با روش‌های پیشنهادی تحقیقات [۱۹] و [۲۰]

الگوریتم	RMSE ($\mu g/m^3$)
پیشنهادی این تحقیق	۱.۸۱۵
پیشنهادی تحقیق [۱۹]	۳.۰۱
پیشنهادی تحقیق [۲۰]	۲.۳۵

همانطور که در جدول ۷ مشخص است روش پیشنهادی این تحقیق توانسته با استفاده از بهینه‌سازی فواصل نامنظم خوشه‌بندی دقتی بهتر از تحقیقات [۱۹] و [۲۰] بدست آورد.

- Y. Hao, H. Flowers, M. M. Monti, and J. R. Qualters, "US census unit population exposures to ambient air pollutants," *International journal of health geographics*, vol. 11, no. 1, p. 3, 2012.
- P. J. Mouser and D. M. Rizzo, "Evaluation of geostatistics for combined hydrochemistry and microbial community fingerprinting at a waste disposal site," in *Critical Transitions in Water and Environmental Resources Management*, 2004, pp. 1–11.
- C. V. Deutsch *et al.*, *JOURNAL. AG GSLIB Geostatistical Software Library and User's Guide*: Oxford University Press, 1998.
- Ilan Levy and D. M. Broday, "Improving modeled air pollution concentration maps by residual interpolation," *Science of the Total Environment*, vol. 598, pp. 780–788, 2017.
- J. Li and A. D. Heap, "A review of comparative studies of spatial interpolation methods in environmental sciences: Performance and impact factors," *Ecological Informatics*, vol. 6, no. 3–4, pp. 228–241, Jul. 2011.
- L. D. Mercer *et al.*, "Comparing universal kriging and land-use regression for predicting concentrations of gaseous oxides of nitrogen (NO_x) for the Multi-Ethnic Study of Atherosclerosis and Air Pollution (MESA Air)," *Atmos Environ*, vol. 45, no. 26, pp. 4412–4420, Aug. 2011.
- حبیبی، ر.، آل شیخ، ع. (۱۳۹۶) ارزیابی کیفیت سروس های درون یابی مکانی در توسعه ی سامانه های پایش آلودگی هوا بر مبنای اینترنت اشیا. نشریه علمی پژوهشی علوم و فنون نقشه برداری. ۶ (۴) ۱۵–۱
- میری، م.، قانعیان، م. ت.، قلیزاده، ع.، یزدانی اول، م.، و نیکونهاد، ع. (۱۳۹۶). تحلیل و پهنه بندی آلودگی هوا شهر مشهد با استفاده از مدل های مختلف تحلیل فضایی. مجله مهندسی بهداشت محیط، ۳(۲)، ۱۴۳–۱۵۴.
- بهاری، ر.، ا.، عباس پور، ر. ع.، و پهلوانی، پ. (۱۳۹۴). پهنه بندی آلودگی ذرات معلق با استفاده از مدل های آماری محلی در GIS (مطالعه موردی، شهر تهران). علوم و فنون نقشه برداری، ۵(۳)، ۱۶۵–۱۷۳.
- A. Eslami and S. M. Ghasemi, "Determination of the best interpolation method in estimating the concentration of environmental air pollutants in Tehran city in 2015," *Journal of Air Pollution and Health*, vol. 3, no. 4, pp. 187–198, 2018.
- A. Eslami and S. M. Ghasemi, "Determination of the best interpolation method in estimating the concentration of environmental air pollutants in Tehran city in 2015," *Journal of Air Pollution and Health*, vol. 3, no. 4, pp. 187–198, 2018.
- Z. Qi, T. Wang, G. Song, W. Hu, X. Li, and Z. Zhang, "Deep air learning: Interpolation, prediction, and feature analysis of fine-grained air quality," *IEEE Transactions on Knowledge and Data Engineering*, vol. 30, no. 12, pp. 2285–2297, 2018.
- S. O. Nabavi, L. Haimberger, and E. Abbasi, "Assessing PM_{2.5} concentrations in Tehran, Iran, from space using MAIAC, deep blue, and dark target AOD and machine learning algorithms," *Atmospheric Pollution Research*, vol. 10, no. 3, pp. 889–903, 2019.
- D. Núñez-Alonso, L. V. Pérez-Arribas, S. Manzoor, and J. O. Cáceres, "Statistical Tools for Air Pollution Assessment: Multivariate and Spatial Analysis Studies in the Madrid Region," *Journal of analytical methods in chemistry*, vol. 2019, 2019.
- R. Shad, M. S. Mesgari, and A. Shad, "Predicting air pollution using fuzzy genetic linear membership kriging in GIS," *Computers, environment and urban systems*, vol. 33, no. 6, pp. 472–481, 2009.
- D. J. Toal, N. W. Bressloff, A. J. Keane, and C. M. E. Holden, "The development of a hybridized particle swarm for kriging hyperparameter tuning," *Engineering optimization*, vol. 43, no. 6, pp. 675–699, 2011.
- R. Webster and M. A. Oliver, *Geostatistics for environmental scientists*. John Wiley & Sons, 2007.
- [6] در این تحقیق از داده های روزانه آلاینده ها استفاده شده است، لذا اگر دقت مدل پیشنهادی برای داده های ساعتی هم بررسی شود می تواند به صورت چشمگیری نتایج تحقیق را بهبود بخشد.
- [7] در مدلسازی آلودگی هوا پارامترهای کاربری های زمین و پارامترهای هواشناسی تاثیر به سزایی دارند، لذا اگر مدلی پیشنهاد شود که این پارامترها را در نظر بگیرد دقت برآورد میزان آلودگی هوا بهبود قابل ملاحظه ای خواهد یافت.
- [8] در این تحقیق به ازای هر مجموعه داده، ترکیب F متفاوتی استفاده شده است، لذا اگر مدلی بیان شود که این تغییرات حذف شود، می تواند مدل پیشنهادی را بهبود بخشد.
- [9] **۶- تقدیر و تشکر**
- از شرکت کنترل کیفیت هوای شهر تهران و سازمان حفاظت محیط زیست جهت ارائه داده های آلودگی هوا مورد نیاز تقدیر و تشکر می گردد.
- انجام این تحقیق با حمایت مالی مرکز مطالعات و همکاری های علمی و بین المللی وزارت علوم، تحقیقات و فناوری با شماره قرار داد ۱۰۸ از نویسنده اول مقاله و همچنین گزنت شماره ۵۳۰۹۳۴۰ معاونت محترم پژوهشی دانشگاه تهران تخصیص یافته به نویسنده اول مقاله انجام شده است که بدینوسیله از مرکز مطالعات و همکاری های علمی بین المللی وزارت علوم و تحقیقات و فناوری و معاونت محترم پژوهشی دانشگاه تهران تقدیر و تشکر می گردد.
- [10] **۷- مشارکت ها**
- مفهوم سازی: م.ر.د، ا.غ و ک.ف، تهیه اطلاعات: م.ر.د، ا.غ و ک.ف، تجزیه و تحلیل: م.ر.د، ا.غ، اخذ بودجه: م.ر.د، روش پیشنهادی: م.ر.د، ا.غ، مدیریت پروژه: م.ر.د، منابع: ا.ه.ا، ارزیابی: م.ر.د، ا.غ، پیاده سازی: ا.غ، نوشتن پیش نویس اولیه: ا.غ و م.ر.د، بررسی و ویرایش: م.ر.د، غ.ر.ش، ی.ر و غ.ر.ن.
- [11] **۸- منابع**
- [1] K. Teather, N. Hogan, K. Critchley, M. Gibson, S. Craig, and J. Hill, "Examining the links between air quality, climate change and respiratory health in Qatar," *Avicenna*, p. 9, 2013.
- [2] R. W. Boubel, D. Vallero, D. L. Fox, B. Turner, and A. C. Stern, *Fundamentals of air pollution*. Elsevier, 2013.
- [3] M. Dehghani, Y. Kamali, M. Ghanbarian, and N. A. Shamsedini, "Study of the Relationship between Indoor," *Outdoor Particle*, 2015.
- [4] A. Zhang, Q. Qi, L. Jiang, F. Zhou, and J. Wang, "Population exposure to PM_{2.5} in the urban area of Beijing," *PloS one*, vol. 8, no. 5, p. e63486, 2013.
- [5] N. W. Jorgensen *et al.*, "Vascular Responses to Long-and Short-Term Exposure to Fine Particulate Matter," *Journal of the American College of Cardiology*, vol. 60, no. 21, pp. 2158–2166, 2012.

- S. Srivastava and I. N. Sinha, "Classification of air pollution dispersion models: a critical review," in *Proceedings of National Seminar on Environmental Engineering with special emphasis on Mining Environment, NSEEME-2004*, 2004. [30]
- E. J. Pebesma, "Multivariable geostatistics in S: the gstat package," *Computers & Geosciences*, vol. 30, no. 7, pp. 683-691, 2004. [31]
- M. Hasanlou and F. Samadzadegan, "ICA/PCA base genetically band selection for classification of Hyperspectral images," in *31st Asian Conference on Remote Sensing*, 2011. [32]
- D. E. Goldberg and J. H. Holland, "Genetic algorithms and machine learning," *Machine learning*, vol. 3, no. 2, pp. 95-99, 1988. [33]
- J. Kennedy, "The particle swarm: social adaptation of knowledge," in *Proceedings of 1997 IEEE International Conference on Evolutionary Computation (ICEC'97)*, 1997, pp. 303-308. [34]
- L. Dao, L. Morrison, G. Kiely, and C. Zhang, "Spatial distribution of potentially bioavailable metals in surface soils of a contaminated sports ground in Galway, Ireland," *Environmental geochemistry and health*, vol. 35, no. 2, pp. 227-238, 2013. [22]
- زارع خوش اقبال, م. غضبان, ف. شریفی, ف. و خسروتهرانی, خ. (۱۳۹۰) استفاده از زمین آمار و GIS در پهنه بندی آلودگی فلزات سنگین در رسوبات تالاب انزلی. فصلنامه زمین. ۶(۱۹), ۳۳-۴۹. [23]
- H. Wackernagel, *Multivariate geostatistics: an introduction with applications*. Springer Science & Business Media, 2013. [24]
- P. A. Burrough and R. A. McDonnell, "Principles of GIS," *Oxford University Press: London, UK*, 1998. [25]
- A. Boufassa and M. Armstrong, "Comparison between different kriging estimators," *Mathematical Geology*, vol. 21, no. 3, pp. 331-345, 1989. [26]
- P. Goovaerts, "Geostatistics for natural resources evaluation. Oxford Univ. Press, New York.," *Geostatistics for natural resources evaluation. Oxford Univ. Press, New York.*, 1997. [27]
- I. Clark and W. V. Harper, *Practical Geostatistics 2000. EcoSSE North America Llc, Ohio*. 2001. [28]
- A. Paz-González, S. Rosa Vieira, and M. Ulloa Guitián, "Spatial variability of nutrient status and soil general properties at different scales," 2003. [29]

زیر نویس ها:

Particle swarm optimization	۲۴	Geostatistics	۱
Best Linear Unbiased Estimator	۲۵	Variogram	۲
Simple Kriging	۲۶	Part Per Billion	۳
Range	۲۷	Ordinary Kriging	۴
Semivariogram	۲۸	Inverse Distance Weighting	۵
Non-bias	۲۹	Land Use Regression	۶
Ordinary Kriging	۳۰	Universal Kriging	۷
Sill	۳۱	Empirical Bayesian Kriging	۸
Range of influence	۳۲	Global Polynomial Interpolation	۹
Nugget effect	۳۳	Universal Kriging	۱۰
Nugget	۳۴	Geographically Weighted Regression	۱۱
Exponential	۳۵	Local Polynomial Interpolation	۱۲
Spherical	۳۶	Radial basis function	۱۳
Gaussian	۳۷	Deep Air Learning	۱۴
Linear	۳۸	Logistic Regression	۱۵
Power model	۳۹	Laplacian Regression	۱۶
Fitness function	۴۰	Recurrent Neural Network	۱۷
Number of Cluster	۴۱	Autoregressive Integrated Moving Average	۱۸
Cross-Validation	۴۲	Dark Target and Deep Blue	۱۹
Root-Mean-Square Error(RMSE)	۴۳	Aerosol Optical Depth	۲۰
May	۴۴	Multi-Angle Implementation of Atmospheric Correction	۲۱
Houston	۴۵	Machine Learning Algorithms	۲۲
		principal component analysis	۲۳