A study on the most important factors affecting the concentration of particulate matter smaller than 10 microns (PM10) using principal component regression

ABSTRACT

Background & objeftive: Air particulate matters which have natural and human made origins have significant effects on the climate, the environment and human health. Several epidemiological studies have shown a direct relationship among the concentrations of suspended particles with different adverse health effects. The general purpose of this research was to determine the most important parameters affecting on the concentration of PM10 in Tajrish station (Tehran) and develop an estimator model for PM10.

Materials & Methods: In this study, a model is constructed using principal component regression (PCR) for the relationship between the hourly concentration of particulate matter smaller than 10 microns with meteorological parameters (WD, WS, T, P, H) and air pollution parameters (CO, NO2, SO2, NOx, NMHC, THC) in Tajrish station (Tehran). The results of the performance evaluation of PCR model were measured in training and testing stages using RMSE, MAE, R and IA as statistical indicators.

Results: The results of principal components import into multiple regression model showed that the most important variable affecting on the concentrations of PM10, are air temperature and the wind speed. Also, CO and SO2 emissions were known as synergic factors for PM10 concentration. The results showed that PCR model is able to estimate 41% of PM10 concentrations in the testing.

Conclusion: The principal components regression analysis showed that meteorological parameters are one of most important factors affecting on the reduction of PM10 concentration in Tajrish station (Tehran).

Document Type: Research article

Keywords: Particulate matter smaller than 10 microns, multiple linear regression, principal component analysis, air pollution, Tehran.

▶ **Citation:** Ehsanzadeh A, Nejadkoorki F, Khodadoostan S. A study on the most important factors affecting the concentration of particulate matter smaller than 10microns (PM10) using principal component regression. *Iranian Journal of Research in Environmental Health*.Summer 2016;2 (2) : 154-164.

Alireza Ehsanzadeh
* M.Sc. Department of Environmental Engineering, Yazd University, Yazd, Iran. (Corresponding Author)
Email: ehsanzadehalireza@gmail.com -

Farhad Nejadkoorki
Assistant professor, Department of Environmental Engineering, Yazd University, Yazd, Iran.
Email: f.nejadkoorki@gmail.com

Sattar Khodadoostan
M.Sc. Department of Statistics, Faculty of Basic Sciences, Yazd University, Yazd, Iran.
sattar.kh.69@gmail.com

Received: 6 February 2016
Accepted: 17 May 2016

بررسی مهم‌ترین عوامل مؤثر بر غلظت ذرات معلق کوچک‌تر از ۱۰ میکرون (PM۱۰) با استفاده از رگرسیون مؤلفه‌های اصلی

چکیده

زمینه و هدف: ذرات معلق موجود در هوا با منشأ طبیعی و انسانی، تأثیرات قابل توجهی بر آب‌وهوا، محیط زیست و سلامت انسان دارند. مطالعات اپیدمیولوژیک متعددی نشان داده‌اند که بین غلظت ذرات معلق با نتایج نامطلوب بهداشتی مختلف ارتباط مستقیمی وجود دارد، لذا مطالعه حاضر با هدف کلی تعیین مهم‌ترین پارامترهای تأثیرگذار بر غلظت $PM_{۱۰}$ ایستگاه تجریش تهران و ایجاد مدل برآوردگر $PM_{۱۰}$ انجام شد.

مواد و روش‌ها: در مطالعه حاضر یک مدل با استفاده از رگرسیون مؤلفه‌های اصلی (PCR) برای بررسی ارتباط بین غلظت ساعتی ذرات معلق کوچک‌تر از ۱۰ میکرون با پارامترهای هواشناسی (سرعت و جهت باد، فشار، رطوبت و دمای هوا) و آلودگی هوای ($NO_۲$، CO، $SO_۲$، NMHC، NOx و THC) ایستگاه تجریش شهر تهران مربوط به دوره زمانی ۱۳۸۵ تا ۱۳۹۰ ارائه شد. نتایج ارزیابی عملکرد مدل PCR در مرحله آموزش و آزمون با استفاده از شاخص‌های آماری R، MAE، RMSE و IA مورد سنجش قرار گرفت.

یافته‌ها: نتایج ورود مؤلفه‌های اصلی به مدل رگرسیون چندگانه نشان داد که مهم‌ترین متغیر مؤثر بر غلظت $PM_{۱۰}$، دمای هوا و سرعت باد می‌باشند. همچنین آلاینده‌های CO و $SO_۲$ عوامل تشدید کننده $PM_{۱۰}$ هستند. نتایج نشان داد مدل PCR در مرحله آزمون قابلیت تخمین ۴۱ درصد مقادیر $PM_{۱۰}$ را دارد.

نتیجه‌گیری: نتایج تحلیل رگرسیون مؤلفه‌های اصلی نشان داد که پارامترهای هواشناسی از عوامل مؤثر بر کاهش غلظت $PM_{۱۰}$ در محدوده ایستگاه تجریش می‌باشند.

نوع مقاله: مقاله پژوهشی

کلیدواژه‌ها: آلودگی هوا، تحلیل مؤلفه‌های اصلی، ذرات با قطر کمتر از ۱۰ میکرون، رگرسیون خطی چندگانه، شهر تهران.

◀ **استناد:** احسان‌زاده ع، نژادکورکی ف، خدادوستان س. بررسی مهم‌ترین عوامل مؤثر بر غلظت ذرات معلق کوچک‌تر از ۱۰ میکرون ($PM_{۱۰}$) با استفاده از رگرسیون مؤلفه‌های اصلی. *فصلنامه پژوهش در بهداشت محیط*. تابستان ۱۳۹۵؛۲ (۲): ۱۵۴-۱۶۴.

علیرضا احسان‌زاده
* کارشناس ارشد، گروه مهندسی محیط زیست، دانشکده منابع طبیعی و محیط زیست، دانشگاه یزد، یزد، ایران. (نویسنده مسئول)
Email: ehsanzadehalireza@gmail.coml

فرهاد نژادکورکی
دانشیار، گروه مهندسی محیط زیست، دانشکده منابع طبیعی و محیط زیست، دانشگاه یزد، یزد، ایران.
Email: f.nejadkoorki@gmail.com

ستار خدادوستان
کارشناس ارشد، گروه آمار، دانشکده علوم پایه، دانشگاه یزد، یزد، ایران.
sattar.kh.69@gmail.com

تاریخ دریافت: ۱۳۹۴/۱۱/۱۷
تاریخ پذیرش: ۱۳۹۵/۰۲/۲۸

مقدمه

آلودگی هوای شهری یک نگرانی روزافزون جهانی است، زیرا اثرات مهمی بر محیط زیست، اقلیم و سلامتی جامعه دارد (۱). ذرات معلق موجود در هوا با منشأ طبیعی و انسانی، تأثیرات قابل توجهی بر آب و هوا، محیط زیست و سلامت انسان دارند. مطالعات اپیدمیولوژیک متعددی نشان داده‌اند که ارتباط مستقیمی بین غلظت ذرات معلق با نتایج نامطلوب بهداشتی مختلف وجود دارد (۲). در کشورهای در حال توسعه، شهرهای بزرگ از جمله شهر تهران با مشکل جدی آلودگی هوا مواجه هستند. امروزه بررسی عوامل مؤثر بر آلودگی هوا به ویژه ذرات معلق، در نواحی شهری به دلیل تأثیر آن بر سلامتی انسان، یکی از موضوعات مهم در تحقیقات زیست‌محیطی است. از این رو بررسی عوامل مؤثر بر غلظت ذرات معلق و تعیین عناصر مؤثر بر آن، از ضروری‌ترین برنامه‌های زیست محیطی در کلان شهرها به‌شمار می‌رود (۳). بررسی عوامل مؤثر بر غلظت آلاینده‌های شاخص در مناطق مختلف شهرهای آلوده، یکی از اقدامات اولیه و گامی مؤثر در ایجاد تصمیمات لازم برای مقابله با آلودگی هواست (۴). امروزه با پیشرفت سریع علم و فناوری، روش‌های آماری و محاسباتی از اهمیت روزافزونی برخوردار شده است. امروزه روش‌های آماری به منظور پیش‌بینی، تخمین و بررسی ارتباط بین متغیرهای مستقل و وابسته مورد استفاده قرار می‌گیرند. دو شاخه مهم آن، روش‌های پارامتریک و غیرپارامتریک هستند. تلفیق رگرسیون خطی چندگانه (MLR) با تحلیل مؤلفه‌های اصلی (PCA) از موضوعات تحقیقات جدید است، زیرا با ترکیب این دو روش آماری می‌توان از مزایای هر دو بهره برد (۵). استفاده بهتر از رگرسیون خطی چندگانه، مستلزم ورود متغیرهای مهم و تأثیرگذار مورد استفاده در آن می‌باشد. برای تعیین بهترین مقادیر پارامترهای رگرسیون خطی چندگانه به طور معمول از روش‌های کاهش بُعد داده‌ها نظیر PCA استفاده می‌شود (۶). به همین منظور در مطالعه حاضر از الگوریتم ترکیبی MLR و PCA به‌عنوان یک روش بهینه‌سازی برای تعیین عوامل مؤثر بر غلظت ساعتی PM_{10} استفاده شد.

بنابراین به دلیل اهمیت این آلاینده بر سلامتی و زیاد بودن مقدار این آلاینده در هوای تهران، آلایندهٔ ذرات معلق کوچک‌تر از ۱۰ میکرون (PM_{10}) و عوامل مؤثر بر آن مورد بررسی قرار گرفت. از جمله مطالعات انجام شده در خصوص اثبات کارایی روش‌های آماری در فرآیند بررسی عوامل مؤثر بر غلظت آلاینده‌های هوا می‌توان به موارد زیر اشاره کرد:

مطالعه واردولوکاس و کاسمومنوس (۲۰۰۸) که با استفاده از آنالیز همبستگی پیرسون و روش تحلیل مؤلفه‌های اصلی به بررسی تأثیر پارامترهای هواشناسی (نظیر سرعت و جهت باد، شدت نور و فشار هوا) و متغیرهای آلودگی هوا (نظیر غلظت SO_2، NO_2 و O_3) بر روی افزایش و کاهش غلظت PM_{10} پرداختند، نشان داد که پارامترهای هواشناسی سرعت و شدت نور از عوامل مؤثر بر کاهش PM_{10} می‌باشند (۶). در مطالعه هاردیلیکوا و همکاران (۲۰۰۸) که با استفاده از رگرسیون خطی تعمیم یافته (GLM) و توزیع گاما به بررسی تأثیر پارامترهای هواشناسی و آلودگی هوا بر غلظت PM_{10} پرداختند، بین پارامترهای هواشناسی در برآورد مقدار PM_{10} ارتباط منفی و معناداری وجود داشت (۷). نتایج مطالعه چن و همکاران (۲۰۱۳) که به پیش‌بینی ذرات معلق کوچک‌تر از ۱۰ میکرون و بررسی اثر آلاینده‌های اصلی هوا و پارامترهای هواشناسی با استفاده از مدل موجک و رگرسیون خطی گام‌به‌گام پرداختند، نشان داد که عوامل هواشناسی، رابطه منفی با غلظت PM_{10} برآوردی دارند (۸). در مطالعه کوا و همکاران (۲۰۰۸) که با کاربرد رگرسیون خطی و شاخص همبستگی پیرسون به بررسی اثر عوامل آلودگی هوا و متغیرهای هواشناسی بر روی غلظت PM_{10} پرداختند، بین غلظت SO_2، NO_2 و CO با مقدار PM_{10} ارتباط افزاینده وجود داشت و سرعت باد از عوامل کاهش دهنده PM_{10} بود (۹). کراسنوا و همکاران (۲۰۱۶) از آنالیز روند تغییرات غلظت PM_{10} و بررسی عوامل مؤثر بر کاهش یا افزایش غلظت آن با استفاده از مدل سری‌های زمانی استفاده و پارامترهای هواشناسی را از عوامل کاهنده غلظت ذرات معلق معرفی کردند (۱۰). هوا و همکاران

(۲۰۱۶) با استفاده از مدل‌های اقتصادسنجی فضایی نظیر مدل اتورگرسیو تأخیر فضایی (SLM) و مدل خطای فضایی (SEM) جهت بررسی عوامل مؤثر بر غلظت ذرات معلق کوچک‌تر از ۲/۵ میکرون ($PM_{2.5}$) اقدام نمودند (۱۱). اونال و همکاران (۲۰۱۱) با استفاده از تحلیل سلسله مراتبی خوشه‌بندی، به بررسی اثر عوامل هواشناسی و آلاینده‌های شاخص هوا بر توزیع فضایی و مکانی ذرات معلق کوچک‌تر از ۱۰ میکرون پرداختند (۱۲). در مطالعه تیان و همکاران (۲۰۱۴) که به بررسی ارتباط بین غلظت PM_{10} با عوامل هواشناسی توسط آنالیز موجک و تجزیه و تحلیل خاکستری پرداختند، بین عوامل هواشناسی با PM_{10} ارتباط منفی و معناداری وجود داشت (۱۳). در مطالعه حاضر با استفاده از رگرسیون خطی چندگانه مبتنی بر تحلیل مؤلفه‌های اصلی جهت بررسی ارتباط بین غلظت ساعتی ذرات معلق کوچک‌تر از ۱۰ میکرون با پارامترهای هواشناسی و آلودگی هوای ایستگاه تجریش شهر تهران ارائه شد. در نهایت مطالعه حاضر با هدف کلی تعیین مهم‌ترین پارامترهای تأثیرگذار بر غلظت PM_{10} و ایجاد مدل برآوردگر ایستگاه تجریش تهران انجام شد.

روش کار

در مطالعه حاضر جهت بررسی تأثیر عوامل مختلف بر غلظت ساعتی PM_{10}، از داده‌های ساعتی غلظت آلاینده‌های هوا و پارامترهای هواشناسی مربوط به ایستگاه تجریش شهر تهران استفاده شد. هدف اولیه استفاده از روش رگرسیون مؤلفه‌های اصلی (PCR)، بررسی ارتباط بین پارامترهای هواشناسی و آلودگی هوا با PM_{10} می‌باشد. در مرحله بعد بر اساس تحلیل انجام شده بر روی داده‌ها، مهم‌ترین عوامل مؤثر بر غلظت PM_{10} با استفاده از روش رگرسیون خطی چندگانه و تحلیل مؤلفه‌های اصلی در نرم‌افزار MATLAB اقدام شد. در این مطالعه از مؤلفه‌های اصلی ایجاد شده با روش PCA به عنوان ورودی مدل رگرسیون در مراحل آموزش و آزمون استفاده شد. همچنین خروجی مدل رگرسیون مقدار PM_{10} در نظر گرفته شد. در این مطالعه ۸۰درصد کل داده‌ها جهت آموزش و ۲۰درصد داده‌ها جهت آزمون مورد استفاده قرار گرفت. تقسیم‌بندی داده‌ها به‌صورت تصادفی و بعد از حذف داده‌های گمشده و پرت انجام گرفت.

داده‌های تحقیق

داده‌های مورد استفاده در این مطالعه به صورت ساعتی و شامل غلظت‌های آلاینده‌های هوا شامل: مونوکسیدکربن (CO)، ذرات معلق کوچک‌تر از ۱۰ میکرون (PM_{10})، دی اکسید نیتروژن (NO_2)، دی‌اکسید گوگرد (SO_2)، اکسیدهای نیتروژن (NOx)، هیدروکربن‌های بدون متان (NMHC) و کل هیدروکربن‌ها (THC) و همچنین برخی پارامترهای هواشناسی نظیر: سرعت باد (WS)، جهت باد (WD)، دمای هوا (T)، فشار (P) و رطوبت هوا (H) بود. داده‌های موجود مربوط به دوره زمانی ۹۰-۱۳۸۵ و شامل ۲۶۷۰۱۳ نمونه ثبت شده بود.

رگرسیون خطی چندگانه

هدف از تحلیل رگرسیون برآورد رابطه‌ای ریاضی است؛ به‌طوری‌که با آن بتوان کمیت متغیری مجهول را با استفاده از متغیرهای معلوم تعیین کرد. رگرسیون چندگانه روشی است که برای بررسی ارتباط خطی بین یک متغیر وابسته و یک یا چند متغیر مستقل استفاده می‌شود (۱۴). مدل ماتریسی رگرسیون چندگانه را می‌توان به صورت معادله زیر نشان داد:

$$Y = X\hat{a} + e \qquad \text{معادله (۱)}$$

که در معادله (۱)، β ماتریس ضرایب رگرسیون، e ماتریس خطای برازش و Y نیز ماتریس پاسخ می‌باشد. با حل معادله بالا بر حسب β خواهیم داشت:

$$\hat{a} = (X'X)^{-1}(X'Y) \qquad \text{معادله (۲)}$$

که در معادله (۳)، X'معکوس ماتریس X است. برای محاسبه معکوس (X'X)، متغیرهای مستقل نباید همبستگی زیادی داشته باشند، زیرا در این صورت ماتریس (X'X) را نمی‌توان معکوس کرد و باعث افزایش خطا در اثر گرد کردن داده‌ها و محاسبات می‌شود. برای رفع این مشکل باید قبل از ساخت مدل رگرسیونی، همبستگی بین متغیرهای مستقل را از بین برد. روش مناسب در

این خصوص، استفاده از تحلیل مؤلفه‌های اصلی روی متغیرهای مستقل ورودی به مدل است. معیار قضاوت برای رفع این مشکل با اجرای تحلیل مؤلفه‌های اصلی روی متغیرهای ورودی، فاکتور تورم واریانس است. عدد ایده‌آل برای فاکتور تورم واریانس ۱ است و مقادیر بزرگ‌تر از ۱۰ برای تورم واریانس، نشانه ناپایداری مدل رگرسیونی است (۱۵، ۱۶).

تحلیل مؤلفه‌های اصلی

تحلیل مؤلفه‌های اصلی، یکی از ارزشمندترین نتایج جبر خطی است، زیرا روش ساده و غیر پارامتری استخراج اطلاعات مربوطه از مجموعه‌های گیج‌کننده است. با این روش متغیرهای مستقل اولیه به مؤلفه‌های جدید و مستقل (بدون همبستگی) تبدیل می‌شوند و سپس از این مؤلفه‌ها به جای متغیرهای اولیه استفاده می‌شود. مؤلفه‌های جدید، ترکیبی خطی از متغیرهای اولیه هستند (۱۷). با استفاده از این روش، ترکیباتی از n متغیر X_n ,...,X_2 ,X_1 برای ایجاد n مؤلفه مستقل PC_n ,...,PC_2 ,PC_1 برقرار می‌شود. در این روش اطلاعات متغیرهای اصلی با کمترین تلفات در مؤلفه‌ها محفوظ می‌ماند. هر مؤلفه اصلی می‌تواند با دنباله زیر مشخص شود:

$$PCi = a_{i1}X_1 + a_{i2}X_2 + \ldots + a_{in}X_n \qquad \text{معادله (۳)}$$

که در معادله (۳)، PC_i معرف مؤلفه مورد نظر، aij بردار ویژه مربوطه و X_i نیز متغیرهای مستقل اولیه است (۱۹). روش کار برای ایجاد مؤلفه‌های اصلی به صورت زیر است:

الف) محاسبه فاکتور KMO[11]

از آنجایی که روش PCA مستلزم وجود و قبول فرضیاتی درباره جامعه مورد مطالعه نیست، از روش‌های ناپارامتری می‌باشد که لازم است امکان استفاده از این روش و نتایج به‌دست آمده از آن به وسیله عامل KMO مشخص شود. مقدار KMO بین صفر تا یک تغییر می‌کند. این عامل با استفاده از ضرایب همبستگی ساده و جزئی طبق معادله (۴) محاسبه می‌شود. در معادله (۴)، rij و aij ضرایب همبستگی ساده و جزئی بین متغیرهای i و j است. با توجه به این معادله، مقادیر بزرگ‌تر KMO مستلزم کوچک بودن ضرایب همبستگی جزئی می‌باشد و بیانگر دقت محاسبات مربوطه، با استفاده از روش PCA است (۱۸).

$$KMO = \frac{\sum_{i=1}^{p}\sum_{i=1}^{p} r_{ij}^2}{\sum_{i=1}^{p}\sum_{i=1}^{p} r_{ij}^2 + \sum_{i=1}^{p}\sum_{i=1}^{p} a_{ij}^2} \qquad i \neq j \qquad (۴)$$

در صورتی که فاکتور KMO بزرگ‌تر از ۰/۵ به‌دست آید، نشان‌دهنده امکان اجرای روش PCA بر روی متغیرهای مستقل است.

ب) استاندارد کردن متغیرهای ورودی:

در این مرحله داده‌های ورودی بر اساس معادله (۵) به نحوی استاندارد می‌شوند که دارای میانگین صفر و انحراف معیار یک باشند.

$$Z = \frac{X - \mu}{\sigma} \qquad (۵)$$

در این معادله، Z معادل مقادیر استاندارد شده داده‌ها، X ماتریس داده‌های ورودی، μ میانگین هر متغیر و σ مقادیر انحراف معیار برای هر متغیر است (۱۹).

ج) محاسبه ماتریس همبستگی (R) برای متغیرهای اولیه:

این ماتریس که ماتریسی متقارن است، میزان تغییرات در نمونه و میزان همبستگی N متغیر را با هم نشان می‌دهد. عضوهای روی قطر اصلی این ماتریس، واریانس متغیرهای ورودی و بقیه درایه‌های این ماتریس، کوواریانس بین متغیرهای ورودی است. از آنجایی‌که برای تشکیل این ماتریس از داده‌های استاندارد شده استفاده شده است، به همین دلیل این ماتریس، معادل ماتریس همبستگی بین متغیرهای ورودی است.

د) محاسبه مقادیر ویژه (λ) و بردارهای ویژه مربوطه از ماتریس همبستگی:

بدین منظور معادله (۶) حل می‌شود:

$$\left|R - \lambda I_p\right| = 0 \qquad (۶)$$

Ip یک ماتریس واحد با بُعد p×p می‌باشد. بنابراین می‌توان p مقدار ویژه مرتب شده $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_p$ را

به‌دست آورد؛ به‌طوری‌که مجموع مقادیر ویژه برابر p باشد. هر مقدار ویژه با اطلاعات مربوط به آن (بردارهای ویژه) ویژگی‌های یک مؤلفه را ارائه می‌دهد. هر مؤلفه نیز درصدی از اطلاعاتی که توسط متغیرهای اولیه بیان می‌شود را در بر می‌گیرد و معادل با بخشی از اطلاعات مسئله است (۱۹). اولین مؤلفه، بیشترین واریانس و آخرین آن، کمترین مقدار واریانس را نشان می‌دهد. انتخاب چند مؤلفه اول که بیشترین مقدار واریانس را دارند، به عنوان مؤلفه‌های اصلی شناخته می‌شوند (۱۷).

۵) اجرای چرخش مناسب روی ماتریس ضرایب مؤلفه‌ها:

چون در تشکیل هر مؤلفه از تمام متغیرهای اولیه استفاده می‌شود، تفسیر مؤلفه‌ها مشکل خواهد بود. به این دلیل روش‌هایی برای تفسیر ساده‌تر مؤلفه‌ها به‌وجود آمده است. این روش‌ها همان چرخش مؤلفه‌ها هستند که به دو نوع چرخش عمودی و مایل تقسیم می‌شوند. به دلیل اینکه در روش چرخش عمودی، استقلال بین مؤلفه‌ها حفظ می‌شود، این نوع چرخش بیشتر مورد استفاده قرار می‌گیرد. در مطالعات علمی بیشتر از چرخش وریماکس استفاده می‌شود که یکی از روش‌های چرخش عمودی است (۱۶، ۱۹).

ارزیابی اعتبار مدل PCR

جهت ارزیابی عملکرد مدل PCR و مقایسه نتایج به‌دست آمده در مراحل آموزش و آزمون از شاخص‌های آماری نظیر شاخص صحت (IA)، جذر میانگین مربعات خطا (RMSE)، میانگین خطای مطلق (MAE) و ضریب همبستگی (R) استفاده می‌شود که معادلات مربوطه به شرح زیر است.

معادله (۷)

$$IA = 1 - \sum_{i=1}^{N}(P_i - O_i)^2 \Big/ \sum_{i=1}^{N}\sum_{i=1}^{N}\left(\left|P_i - \bar{O}\right| + \left|O_i - \bar{O}\right|\right)^2$$

معادله (۸)

$$RMSE = (1/N \sum_{i=1}^{N} (P_i - O_i)^2)^{0.5}$$

معادله (۹)

$$MAE = 1/N \sum_{i=1}^{N} |P_i - O_i|$$

معادله (۱۰)

$$R = \sum_{i=1}^{N}\left((P_i - \bar{O})^2 - \sum_{i=1}^{N}(P_i - O_i)^2\right) \Big/ \sum_{i=1}^{N}(P_i - \bar{O})^2$$

که N تعداد کل داده‌ها، Pi مقادیر پیش‌بینی شده، Oi مقادیر مشاهده یا محاسبه شده و $\bar{O}$ میانگین مقادیر مشاهده یا محاسبه می‌باشند.

یافته‌ها

نتایج حاصل از بررسی آماری داده‌های ایستگاه تجریش در جدول ۱ و در ادامه نتایج حاصل از تحلیل عوامل مؤثر بر غلظت ساعتی PM_{10} در ایستگاه پایش کیفیت هوای تجریش ارائه شده است. در این مطالعه تحلیل و بررسی عوامل مؤثر بر PM_{10} به وسیله روش رگرسیون خطی مبتنی بر تحلیل مؤلفه‌های اصلی انجام شد.

جدول ۱. مشخصات آماری هر یک از پارامترهای ایستگاه تجریش

نام	PM_{10}	NO_2	SO_2	CO	NOx	THC	NMHC	H	P	T	WD	WS
واحد	$\mu g/m^3$	ppm	ppm	ppm	ppm	ppm	ppm	RH٪	mBar	DegC	Deg	m/s
حداقل	۰/۰	۰/۰	۰/۰۰۱	۰/۰۰۱	۰/۰۰۱	۰/۰۱۷	۰/۰۱۵	۲/۷۶	۰/۳	۹/۹۷	۰/۰	۰/۰۰۱
حداکثر	۵۱۵	۰/۲۴	۰/۰۵۴	۰/۴۴	۰/۲۹	۰/۹۴	۸/۹۲	۰/۸۵	۰/۸	۰/۲۱	۳۵۵	۱۷/۰۹
میانگین	۰/۵	۰/۰۴	۰/۰۱۷	۴/۰۱	۰/۰۶	۵/۴	۲/۰۱	۳۴/۷	۸۳۳	۱۹/۲	۰/۶۵	۱/۰۵
چولگی	۱/۴	۱/۳۶	۰/۶۴	۰/۸۵	۱/۱۷	۰/۲۳	۰/۹۷	۱/۲۶	۱۶/۷	۰/۳۴	-۰/۲۲	۲/۶۵
کشیدگی	۴	۳/۳۶	۰/۷۶	۰/۳	۱/۴۴	۰/۰۴	۱/۶۵	۰/۹۶	۰/۵۸	۰/۶۶	-۰/۱۱	۳۹/۳

اجرای PCA جهت پیش‌پردازش داده‌ها

بر اساس تحلیل اولیه، بین متغیرهای ورودی مورد استفاده در ایستگاه تجریش همبستگی معنی‌داری وجود داشت که برای تجزیه و تحلیل پارامترهای ورودی به مدل رگرسیون، از روش تحلیل مؤلفه‌های اصلی استفاده شد. همچنین جهت بررسی امکان اجرای تحلیل مؤلفه‌های اصلی از آزمون Kaiser-Meyer-Olkin (KMO)

استفاده شد. مقدار KMO =۰/۵۰۶ امکان اجرای روش PCA را تأیید کرد. برای اجرای این روش، پس از استاندارد کردن متغیرهای ورودی، ماتریس همبستگی پیرسون تشکیل شد که نتایج آن در جدول ۲ ارائه شده است. با حل دستگاه معادله (۶)، ۱۱ مقدار ویژه و به ازای هر مقدار ویژه، ۱۱ بردار ویژه برای ایستگاه تجریش حاصل شد که با استفاده از آن‌ها، مؤلفه‌های اصلی از متغیرهای اولیه به‌دست آمد. مقدار عددی هر مؤلفه با تقسیم مقادیر ویژه به‌دست آمده بر تعداد متغیرهای مورد استفاده به‌دست آمد. درصد پراکندگی نیز از تقسیم مقدار عددی هر مؤلفه بر تعداد متغیرهای مورد استفاده محاسبه شد. مشخصات هر مؤلفه در جدول ۳ آورده شده است. مؤلفه‌هایی که مقادیر ویژه آن‌ها بیشتر از یک بود، به عنوان مؤلفه‌های اصلی انتخاب شدند. اسکری پلات مؤلفه‌های اصلی نمایش در شکل ۱ داده شده است. با توجه به جدول ۳، مقادیر ویژه مؤلفه اول تا پنجم در ایستگاه تجریش بزرگ‌تر از ۱ بوده و به صورت جمعی دارای واریانس تجمعی ۷۸/۳۲۷ درصد کل داده‌ها بودند، همچنین مقدار اولین مؤلفه برابر ۲/۸۵۴ بود که ۲۵/۹۴۱ درصد از کل واریانس موجود در سری داده‌ها را توجیه می‌کند. بدین ترتیب مؤلفه اول تا پنجم در ایستگاه تجریش دارای مقادیر ویژه بزرگ‌تر از یک بودند که می‌توان این مؤلفه‌ها را به عنوان مؤلفه‌های اصلی انتخاب و به‌عنوان ورودی مدل رگرسیون استفاده کرد.

جدول ۲. ماتریس همبستگی پیرسون بین متغیرهای ورودی روش PCA

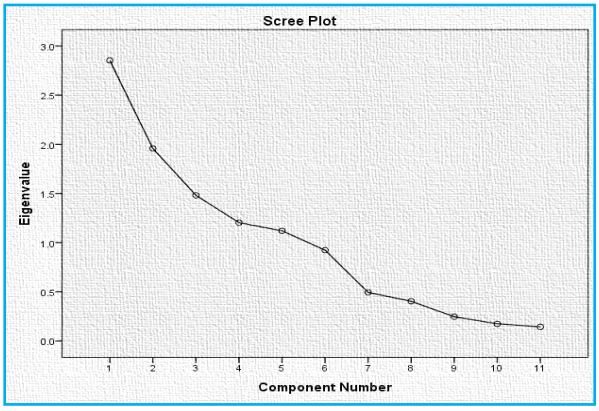
**	PM_{10}										
PM_{10}	۱	NO_2									
NO_2	۰/۲۴	۱	SO_2								
SO_2	۰/۲۹	۰/۲۹	۱	CO							
CO	۰/۳۰	۰/۵۰	۰/۴۴	۱	NOx						
NOx	۰/۲۶	۰/۸۷	۰/۳۸	۰/۶۲	۱	THC					
THC	۰/۱۳	۰/۱۲	-۰/۰۵	۰/۲۴	۰/۱۲	۱	NMHC				
NMHC	۰/۲۲	۰/۲۳	۰/۰۰	۰/۵۲	۰/۲۵	۰/۸۵	۱	HUM			
HUM	-۰/۰۲	۰/۱۸	۰/۲۵	۰/۲۲	۰/۲۸	۰/۱۲	۰/۰۲	۱	Press		
Press	۰/۰۵	۰/۰۸	۰/۰۳	۰/۰۵	۰/۰۸	۰/۱۰	۰/۰۷	۰/۱۰	۱	Temp	
Temp	-۰/۰۱	-۰/۱۹	-۰/۲۱	-۰/۲۲	-۰/۲۷	-۰/۱۹	-۰/۰۸	-۰/۶۵	-۰/۰۲	۱	WD
WD	-۰/۰۸	-۰/۱۰	-۰/۱۳	-۰/۲۰	-۰/۱۲	۰/۰۰	-۰/۱۱	-۰/۰۹	۰/۱۴	۰/۱۴	۱	WS
WS	۰/۰۶	-۰/۱۰	-۰/۱۱	-۰/۱۴	-۰/۱۸	-۰/۲۰	-۰/۰۹	-۰/۲۷	۰/۰۱	۰/۱۹	-۰/۰۵	۱

جدول۳. مشخصات مؤلفه‌های ایجاد شده با روش PCA

	ایستگاه تجریش		
مؤلفه‌ها	مقادیر ویژه	درصد واریانس	واریانس تجمعی
۱	۲/۸۵۴	۲۵/۹۴۱	۲۵/۹۴۱
۲	۱/۹۵۸	۱۷/۷۹۶	۴۳/۷۳۸
۳	۱/۴۸۱	۱۳/۴۶۸	۵۷/۲۰۶
۴	۱/۲۰۳	۱۰/۹۳۵	۶۸/۱۴۱
۵	۱/۱۲۰	۱۰/۱۸۶	۷۸/۳۲۷
۶	۰/۹۲۳	۸/۳۹۱	۸۶/۷۱۹
۷	۰/۴۹۴	۴/۴۹۰	۹۱/۲۰۹
۸	۰/۴۰۴	۳/۶۷۰	۹۴/۸۷۹
۹	۰/۲۴۶	۲/۲۳۹	۹۷/۱۱۸
۱۰	۰/۱۷۴	۱/۵۷۹	۹۸/۶۹۸
۱۱	۰/۱۴۳	۱/۳۰۲	۱۰۰/۰۰۰

ضرایب بردارهای ویژه ایستگاه تجریش در جدول ۴ آمده است. بر اساس جدول ۴، در مؤلفه اول ضرایب مربوط به متغیرهای THC و NMHC دارای مقادیر بیشتری بودند و تأثیر

بسزایی در تشکیل این مؤلفه داشتند که با قلم پررنگ مشخص شده است. در مؤلفه دوم متغیرهای NO_2 و NOx و در مؤلفه سوم، مقدار فشار هوا و جهت باد، عوامل اصلی ایجاد این مؤلفه بودند. همچنین در تشکیل مؤلفه چهارم متغیرهای CO_2، SO_2 و رطوبت هوا و در نهایت در مؤلفه پنجم، دمای هوا و جهت باد، مؤثرترین عوامل ایجاد این مؤلفه بودند. معادله‌های (۱۱) تا (۱۵) مؤلفه‌های اصلی ایجاد شده با متغیرهای مهم جهت ورود به تحلیل رگرسیون را نشان داده‌اند.



شکل ۱. اسکری پلات مؤلفه‌های اصلی بر اساس مقادیر ویژه

جدول ۴. ضرایب بردارهای ویژه برای ایجاد مؤلفه اصلی

مؤلفه‌های اصلی					
ایستگاه تجریش					
پارامترها	۱	۲	۳	۴	۵
NO_2	۰/۰۵۴	۰/۹۵۲	۰/۰۲۱	۰/۰۴۱	۰/۰۱۷-
SO_2	۰/۰۳۷	۰/۱۱۲	۰/۱۴۵-	۰/۷۴۲	۰/۰۸۱-
CO	۰/۴۳۳	۰/۳۰۸	۰/۱۴۴-	۰/۵۸۱	۰/۰۴۸-
NOx	۰/۰۹۲	۰/۹۲۵	۰/۰۰۷	۰/۲۳۴	۰/۰۴۸-
THC	۰/۹۱۱	۰/۰۵۴	۰/۰۲۱-	۰/۰۴۷	۰/۰۷۰-
NMHC	۰/۹۲۲	۰/۰۵۹	۰/۰۵۳-	۰/۰۹۳	۰/۰۰۶-
H	۰/۰۲۵	۰/۰۵۰	۰/۳۷۰	۰/۷۰۷	۰/۰۹۹
P	۰/۰۷۵-	۰/۰۴۵	۰/۸۶۹	۰/۱۴۲	۰/۰۴۵
T	۰/۰۹۰-	۰/۰۵۴-	۰/۲۰۶	۰/۳۱۶-	۰/۷۹۵
WD	۰/۰۲۲-	۰/۰۳۱-	۰/۸۳۸	۰/۱۴۸-	۰/۱۱۹
WS	۰/۰۰۸-	۰/۰۱۷-	۰/۰۰۹	۰/۲۰۶	۰/۹۱۷

معادله (۱۱)

$$PC_1 = (0/911 \times THC) + (0/922 \times NMHC)$$

معادله (۱۲)

$$PC_2 = (0/952 \times NO_2) + (0/925 \times NOx)$$

معادله (۱۳)

$$PC_3 = (0/869 \times P) + (0/838 \times WD)$$

معادله (۱۴)

$$PC_4 = (0/742 \times SO_2) + (0/571 \times CO) + (0/707 \times H)$$

معادله (۱۵)

$$PC_5 = (0/795 \times T) + (0/917 \times WS)$$

مدل رگرسیون مبتنی بر مؤلفه‌های اصلی

به طوری کلی در مطالعه حاضر از روش PCA به منظور حذف همبستگی بین عوامل ورودی روش رگرسیون و تحلیل آسان‌تر عوامل مؤثر بر PM_{10} استفاده شد. به منظور توسعه مدل رگرسیون مؤلفه‌های اصلی از ۵ مؤلفه ایجاد شده ایستگاه تجریش به عنوان ورودی مدل رگرسیون استفاده شد. به منظور بررسی پیش‌فرض نرمال بودن توزیع خطای باقی‌مانده‌های مدل رگرسیون، از آزمون دوربین – واتسون استفاده شد. در صورتی که مقدار این آماره بین ۱/۵-۲/۵ باشد، نشان‌دهنده نرمال بودن توزیع خطا است. در مدل رگرسیون ایجاد شده مقدار این آماره ۱/۷ به‌دست آمد که نشان از نرمال بودن توزیع خطا داشت. همچنین جهت بررسی پیش‌فرض عدم هم‌خطی (استقلال خطاها) از آماره تورم واریانس استفاده شد که مقدار ایده‌آل آن ۱ و مقادیر بیشتر از ۱۰ نشان‌دهنده مشکل‌ساز بودن هم‌خطی بین متغیرهای مستقل در مدل رگرسیون است. مشکل همبستگی بین متغیرهای مستقل نیز با استفاده از روش PCA رفع شد. پس از تأیید نرمال بودن توزیع مقادیر خطا و رفع مشکل همبستگی در متغیرهای مستقل، مدلی مناسب با استفاده از روش رگرسیون خطی چندگانه برای تعیین مهم‌ترین متغیرهای مؤثر بر غلظت PM_{10} مقدار شاخص کیفیت هوا بسط یافت. نتایج ورود مؤلفه‌های اصلی به مدل رگرسیون ایستگاه تجریش در جدول ۵ آمده است. از پنج مؤلفه ایجاد شده به عنوان ورودی مدل رگرسیون استفاده شد. در معادله (۱۶) مدل رگرسیون مؤلفه‌های اصلی برآوردگر PM_{10} و ضرایب هر یک از مؤلفه‌های اصلی ارائه شده است. نتایج ارزیابی عملکرد مدل PCR برآوردگر PM_{10} در

مرحله آموزش و آزمون در جدول ۶ نشان داده شده است. در شکل ۲ پلات‌های مربوط به مقادیر واقعی و برآورد شده در مرحله آموزش و آزمون مدل PCR نشان داده شده است.

معادله (۱۶)

$$PM_{10}= 117/071+(5/851\times PC_1)+(10/853\times PC_2)+(11/302\times PC_3)+(32/630\times PC_4)+(52/872\times PC_5)$$

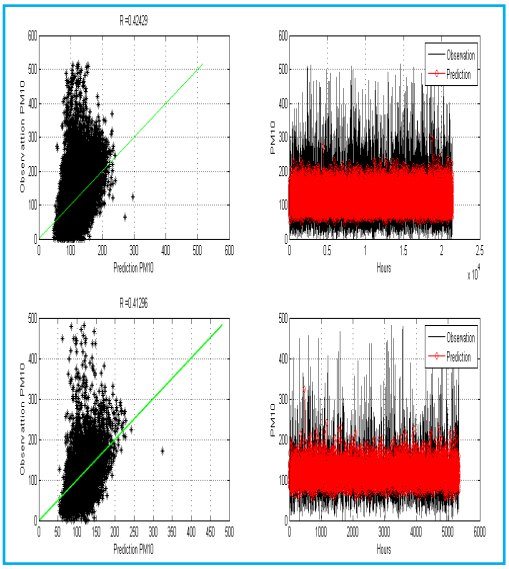
جدول ۵. نتایج ورود مؤلفه‌های اصلی به مدل رگرسیون چندگانه ایستگاه تجریش

مؤلفه‌ها	ضرایب	فاکتور تورم واریانس	سطح معنی‌داری
(Constant)	۱۱۷/۰۷۱	———	۰/۰۰۰
PC_1	۵/۸۵۱	۱/۰۰۰	۰/۰۰۰
PC_2	۱۰/۸۳۵	۱/۰۰۰	۰/۰۰۰
PC_3	۱۱/۳۰۲	۱/۰۰۰	۰/۰۰۰
PC_4	۳۲/۶۳۰	۱/۰۰۰	۰/۰۰۰
PC_5	۵۲/۸۷۲	۱/۰۰۰	۰/۰۰۰

جدول ۶- نتایج ارزیابی مدل رگرسیون مؤلفه‌های اصلی

PCR مدل	IA (1)	R (1)	MAE (0)	RMSE (0)
مرحله آموزش	۱	۰/۴۲۴	۴۲/۷۶	۵۸/۷۹
مرحله آزمون	۱	۰/۴۱۲	۴۳/۰۲	۵۹/۰۵

اعداد داخل پرانتز مقادیر ایده‌آل هر شاخص را نشان می‌دهد.



شکل ۲. پلات‌های مربوط به مرحله آموزش و آزمون مدل PCR

بحث

با توجه به جدول ضرایب مؤلفه‌ها، ضریب عددی برخی متغیرهای ورودی به PCA از وزن بیشتری برخوردار بودند که با قلم پررنگ و رنگ قرمز مشخص شده است. بر اساس نتایج ورود مؤلفه‌های اصلی به مدل رگرسیون چندگانه ایستگاه تجریش، مؤلفه پنجم با ضریب ۵۲/۸۷۲ بیشترین اهمیت را داشت. مهم‌ترین متغیرهای تشکیل دهنده این مؤلفه، دمای هوا و سرعت باد بودند. همچنین مؤلفه چهارم با ۳۲/۶۳۰ بعد از مؤلفه پنجم از اهمیت بسزایی در برآورد PM_{10} برخوردار بود. مؤلفه سوم، دوم و اول به ترتیب با ضرایب ۱۱/۳۰۲، ۱۰/۸۳۵ و ۵/۸۵۱ در رتبه‌های سوم، چهارم و پنجم بودند. نتایج ارزیابی عملکرد مدل PCR در مرحله آموزش و آزمون با استفاده از شاخص‌های آماری R، MAE، RMSE و IA مورد سنجش قرار گرفت. بر اساس نتایج به‌دست آمده، مدل PCR در مرحله آموزش، قابلیت تخمین ۴۲درصد مقادیر PM_{10} را داشت، همچنین در مرحله آزمون نتایج تقریباً مشابهی به‌دست آمد. بر اساس نتایج به‌دست آمده از جدول ۲ ضریب همبستگی پیرسون و معادله (۱۶) می‌توان اینگونه استنباط کرد که مؤلفه پنجم و چهارم بیشترین اهمیت را در برآورد PM_{10} داشتند و بر اساس ضریب همبستگی پیرسون، متغیر سرعت باد دارای رابطه منفی با PM_{10} بود و بر اساس ضریب مؤلفه پنجم در معادله (۱۶)، مهم‌ترین عامل در کاهش غلظت PM_{10} بودند. همچنین متغیرهای دما و رطوبت هوا دارای رابطه منفی با غلظت PM_{10} و از عوامل کاهش‌دهنده PM_{10} موجود در هوا بودند. لازم به ذکر است آلاینده‌های CO و SO_2 از عوامل تشدید کننده غلظت PM_{10} بوده و از اهمیت بسزایی در افزایش غلظت این آلاینده در محدوده ایستگاه پایش کیفیت هوای تجریش شهر تهران برخوردار بودند. نتایج مطالعه حاضر با مطالعه واردولوکاس و کاسمومنوس (۲۰۰۸) که با استفاده از تحلیل مؤلفه‌های اصلی و ضریب همبستگی پیرسون به بررسی تأثیر پارامترهای هواشناسی و آلودگی هوا بر روی غلظت PM_{10} پرداختند و رابطه منفی و کاهنده‌ای بین غلظت سرعت باد و ازن سطح زمین نشان دادند (۶)، و همچنین با نتایج

مطالعه کوا و همکاران (۲۰۰۸) که با استفاده از مدل رگرسیون گام به گام و آنالیز موجک نسبت به بررسی تأثیر پارامترهای هواشناسی و آلودگی هوا بر PM_{10} پرداختند و نشان دادند که با افزایش آلاینده‌های NO_2 و SO_2، مقدار PM_{10} نیز افزایش می‌یابد، همخوانی داشت (۹). همچنین نتایج مطالعه حاضر با مطالعه چن و همکاران (۲۰۱۳) که با استفاده از مدل درخت تصمیم و آنالیز موجک به پیش‌بینی و بررسی عوامل مؤثر بر غلظت PM_{10} پرداختند و نشان دادند که پارامترهای هواشناسی نظیر سرعت باد و دمای هوا از عوامل کاهش دهنده PM_{10} می‌باشند، با این تفاوت که مدل‌های هوش مصنوعی عملکرد مطلقاً برتری در برآورد مقدار PM_{10} نسبت به مدل PCR مورد استفاده در این تحقیق دارند و نتایج تحقیق با مطالعه حاضر همخوانی دارد (۸) در مطالعه تیان و همکاران (۲۰۱۴) که با استفاده از مدل‌های هوش محاسباتی نسبت به برآورد PM_{10} پرداختند و با مقایسه شاخص‌های آماری R، RMSE و MAE نشان دادند که مدل‌های هوش محاسباتی عملکرد مطلقاً برتری نسبت به مدل PCR مورد استفاده در این مطالعه دارد (۱۳).

در این تحقیق، مطالعه در سطح یک ایستگاه سنجش آلودگی هوا انجام شد که به دلیل اهمیت بسیار زیاد موضوع آلودگی هوا ناشی از ذرات معلق در شهرهای آلوده نظیر تهران پیشنهاد می‌شود در مطالعات مشابه آینده مطالعه در سطح وسیع‌تر انجام شود و جهت بررسی دقیق‌تر عوامل مؤثر بر کاهش یا افزایش غلظت PM_{10} پیشنهاد می‌شود پژوهش در سطح گسترده‌تر و از روش‌های مختلف نظیر ترکیب الگوریتم‌های تکاملی مثل ژنتیک و رقابت استعماری با شبکه‌های عصبی مصنوعی نسبت به بررسی عوامل مؤثر بر PM_{10} و توسعه مدل‌های برآوردگر با دقت بسیار بالا اقدام شود. دست‌آوردهای این تحقیق می‌تواند به عنوان یک مطالعه پایه در حوزه آلودگی هوا ناشی از ذرات معلق در شهر تهران مورد استفاده مدیران و برنامه‌ریزان شهری جهت اتخاذ تدابیر کنترلی برای پیشگیری از آلودگی هوای شهری و تعیین عناصر نامطلوب بر کیفیت هوای تنفسی قرار گیرد.

نتیجه گیری

پارامترهای هواشناسی و آلودگی هوا اثرات متفاوتی بر غلظت PM_{10} دارند. نتایج تحلیل رگرسیون مؤلفه‌های اصلی نشان داد که سرعت باد مهم‌ترین عامل در کاهش غلظت PM_{10} در محدوده ایستگاه تجریش می‌باشد. همچنین آلاینده‌های CO و SO_2 از عوامل تشدید کننده غلظت PM_{10} می‌باشند.

تشکر و قدردانی

بدین‌وسیله از سازمان حفاظت محیط زیست به خاطر در اختیار گذاشتن داده‌های تحقیق تشکر و قدردانی می‌شود.

Reference:

1. Li L, Qian J, Ou C-Q, Zhou Y-X, Guo C, Guo Y. Spatial and temporal analysis of Air Pollution Index and its timescale-dependent relationship with meteorological factors in Guangzhou, China, 2001–2011. Environmental Pollution. 2014;190:75-81.
2. You W, Zang Z, Zhang L, Li Z, Chen D, Zhang G. Estimating ground-level PM10 concentration in northwestern China using geographically weighted regression based on satellite AOD combined with CALIPSO and MODIS fire count. Remote Sensing of Environment. 2015;168:276-85.
3. Petelin D, Grancharova A, Kocijan J. Evolving Gaussian process models for prediction of ozone concentration in the air. Simulation modelling practice and theory. 2013;33:68-80.
4. Russo A, Raischel F, Lind PG. Air quality prediction using optimal neural networks with stochastic variables. Atmospheric Environment. 2013;79:822-30.
5. Chai Y, Jia L, Zhang Z, editors. Mamdani model based adaptive neural fuzzy inference system and its application in traffic level of service evaluation. Fuzzy Systems and Knowledge Discovery, 2009 FSKD'09 Sixth International Conference on; 2009: IEEE.
6. Vardoulakis S, Kassomenos P. Sources and factors affecting PM 10 levels in two European cities: implications for local air quality management. Atmospheric Environment. 2008;42(17):3949-63.
7. Hrdličková Z, Michalek J, Kolář M, Veselý V. Identification of factors affecting air pollution by dust aerosol PM 10 in Brno City, Czech Republic. Atmospheric Environment.

2008;42(37):8661-73.
8. Chen Y, Shi R, Shu S, Gao W. Ensemble and enhanced PM 10 concentration forecast model based on stepwise regression and wavelet analysis. Atmospheric Environment. 2013;74:346-59.
9. Kuo C-Y, Chen P-T, Lin Y-C, Lin C-Y, Chen H-H, Shih J-F. Factors affecting the concentrations of PM 10 in central Taiwan. Chemosphere. 2008;70(7):1273-9.
10. Krasnov H, Katra I, Friger M. Increase in dust storm related PM 10 concentrations: A time series analysis of 2001–2015. Environmental Pollution. 2016;213:36-42.
11. Hao Y, Liu Y-M. The influential factors of urban PM 2.5 concentrations in China: a spatial econometric analysis. Journal of Cleaner Production. 2016;112:1443-53.
12. Unal YS, Toros H, Deniz A, Incecik S. Influence of meteorological factors and emission sources on spatial and temporal variations of PM10 concentrations in Istanbul metropolitan area. Atmospheric Environment. 2011;45(31):5504-13.
13. Tian G, Qiao Z, Xu X. Characteristics of particulate matter (PM 10) and its relationship with meteorological factors during 2001–2012 in Beijing. Environmental Pollution. 2014;192:266-74.
14. Rawlings JO, Pantula SG, Dickey DA. Applied regression analysis: a research tool: Springer Science & Business Media; 1998.
15. Chatterjee S, Hadi AS. Regression analysis by example: John Wiley & Sons; 2015.
16. Nouril, R.E., K. Ashrafi, and A. Azhdarpour, Comparison of ANN and PCA based multivariate linear regression applied to predict the daily average concentration of CO: A case study of Tehran. 2008. Journal of the Earth and Space Physics.2008:34(2).
17. Çamdevýren H, Demýr N, Kanik A, Keskýn S. Use of principal component scores in multiple linear regression models for prediction of Chlorophyll-a in reservoirs. Ecological Modelling. 2005;181(4):581-9.
18. Abdul-Wahab SA, Bakheit CS, Al-Alawi SM. Principal component and multiple regression analysis in modelling of ground-level ozone and factors affecting its concentrations. Environmental Modelling & Software. 2005;20(10):1263-71.
19. Jolliffe I. Principal component analysis: Wiley Online Library; 2002.