Upgrading the Estimation of Daily PM$_{10}$ Concentrations Utilizing Prediction Variables Reflecting Atmospheric Processes

Konstantinos Dimitriou*

Laboratory of Meteorology, Department of Physics, University of Ioannina, Ioannina 45110, Greece

ABSTRACT

This paper formulates a Multiple Linear Regression Model (MLRM), for the estimation of daily PM$_{10}$ concentrations in background urban areas. 24-hour backward air mass trajectories, NO$_2$ concentrations and gridded (1° × 1° resolution) Aerosol Optical Depth (AOD) observations from MODIS were used in order to compose the model’s predictor variables. As a supplement to local combustion/non-combustion contributions, the suggested method intends to comprise and quantify the effect that transboundary PM sources and wind dispersion on particulate air pollution levels. The proposed technique was implemented at a background sampling site in Birmingham (United Kingdom) and the results were compared with the outcome of a Simple Linear Regression Model (SLRM) which contained only one predictor variable expressing local combustion. Various statistical indices signified the upgraded performance of the MLRM, in comparison with SLRM, thus the participation of long range transport and wind dispersion variables in the MLRM was successful. According to the MLRM’s findings, anthropogenic combustion (traffic, heating) is the strongest source of PM$_{10}$ in the selected background urban area, followed by local non-combustion emissions and long range transport. Extreme PM$_{2.5}$ intrusions from continental Europe also emerged.

Keywords: PM$_{10}$; MODIS; Aerosol Optical Depth; Wind dispersion; Multiple Linear Regression.

INTRODUCTION

The connection of outdoor particulate air pollution with adverse health effects, has been highlighted in multiple recent publications worldwide (Cruz et al., 2015; Dai et al., 2015, Kim et al., 2015; Angelici et al., 2016; Tsangari et al., 2016). Hence, the European Union (EU) has established upper concentration limits for PM$_{10}$ (Particulate Matter with diameter less than 10 µm) and PM$_{2.5}$ (Particulate Matter with diameter less than 2.5 µm), in order to protect the European population from being exposed to excessive levels of inhalable aerosols (Council Directive 1999/30/EC). In the city of Birmingham (United Kingdom), where this study takes place, an important urban background contribution to indoor particulate matter, from penetration of outdoor particles, has also been indicated (Hock et al., 2008; Jones et al., 2000). Thus, sources and factors influencing aerosol levels must be well analyzed and controlled. Particle composition in Birmingham, clearly revealed an urban aerosol profile made up by combustion generated carbonaceous particles, secondary sulphates and nitrates from regional transport and predominantly coarse (PM$_{COARSE}$ = PM$_{10}$ – PM$_{2.5}$) dusts (Harrison et al., 2004; Taiwo, 2016). Non-exhaust particles derived from tire wear and break friction, also constitute an important source of PM in Birmingham, according to Pant et al. (2015); Taiwo et al. (2014) and Harrison et al. (2003).

Apart from local emissions, PM levels in Birmingham are also strongly dependent from the prevailing atmospheric patterns. More specifically, calm conditions attributed to the influence of high pressure anticyclonic systems were linked with poor air quality situations during winter, due to the entrapment of airborne particles (McGregor et al., 1999). Air mass trajectories are a widely used tool for the identification of PM transportation. A cluster analysis of three dimensional (3D) backward trajectories was implemented by Makra et al. (2011, 2013) and Makra et al. (2010, 2016) to indicate potential regional/transboundary source regions of PM$_{10}$ and biogenic material respectively, enriching aerosol levels in several European cities (e.g., Szeged, Bucharest, Helsinki, Thessaloniki and Hamburg). From the air mass trajectory analysis of Vardoulakis and Kassomenos (2008), it was found that the long range transport particles from continental Europe had a marked effect on PM$_{10}$ background levels in Birmingham. In addition, a multivariate linear regression model, developed by Kavouras et al. (2013), based on air mass residence time variables, identified Central/Eastern Europe as the area with highest PM$_{2.5}$ and PM$_{COARSE}$ contribution in Birmingham. The inflow of particles from

* Corresponding author.
Tel.: +30-6976117441
E-mail address: kdimitriou80@gmail.com
inner Europe, due to Eastern atmospheric circulation, was also emerged in other major British cities as London and Edinburgh (Buchanan et al., 2002; Dimitriou and Kassomenos, 2014a; Beddows et al., 2015; Heal et al., 2005).

In this paper, surface air pollution data, backward air mass trajectories and satellite Aerosol Optical Depth (AOD) observations, were combined in a multiple linear regression model, purposing to estimate daily PM$_{10}$ concentrations. The produced formula was implemented in a background urban area of Birmingham. As a supplement to local combustion/non-combustion contributions, the influences of long range transport and wind dispersion were also quantified. The model’s performance was evaluated by various statistical indicators, in order to reveal possible improvements, generated by the participation of variables reflecting atmospheric processes.

DATA AND METHODOLOGY

Data

Air Pollution Data

Air pollution levels in the city of Birmingham (United Kingdom) are systematically monitored by an air quality network which includes 12 ground-based environmental stations. For this work, daily concentration data of PM$_{10}$, PM$_{2.5}$ and NO$_2$ (Table 1) were derived from the urban background station: Birmingham Tyburn (EU code: GB0851A), situated to the North-East (NE) of the city’s center (Longitude: 1.83°W, Latitude: 52.51°N, Altitude: 93 m). The sampling site is situated within a self-contained, air-conditioned housing located within a car park (Fig. 1(a)), whereas the inlets of the receptor are placed at a height of 4 m Above Ground Level (AGL). The type of area around the station is mainly residential and commercial, with some vegetation and no significant industrial facilities (Fig. 1(b)), whereas the nearest road (other than car park access road) is approximately 60 m to the North (A38 Tyburn Road). In addition, the monitoring site is located far away (approximately 600 m) from the M6 motorway. For all these reasons, no direct combustion emissions are expected in the vicinity of the station and thus the requirements for the identification of long range transport impacts are fulfilled. The environmental station is equipped with Tapered Element Oscillating Microbalance (TEOM) samplers and Chemiluminescence sensors for the monitoring of PM and NO$_2$ respectively. All air pollution data, extending from 1/1/2011 to 31/7/2012, were obtained from the website (http://www.eea.europa.eu/themes/air/air-quality/map/airbase) of the EU air quality database.

Satellite Data

Daily AOD observations (1° × 1° resolution grid) at 550 nm, recorded over North–West (NW) Europe during the preceding day of each day included in the studied time interval (1/1/2011–31/7/2012), were also downloaded from the portal Giovanni (http://giovanni.sci.gsfc.nasa.gov) of

| Table 1. Statistics for PM$_{10}$, PM$_{2.5}$, PM$_{COARSE}$ ($=$ PM$_{10}$ – PM$_{2.5}$) and NO$_2$ data. |
|-------------------------------------------------|-----------------|-----------------|-----------------|-----------------|
| Variable             | PM$_{10}$ (μg m$^{-3}$) | PM$_{2.5}$ (μg m$^{-3}$) | PM$_{COARSE}$ (μg m$^{-3}$) | NO$_2$ (μg m$^{-3}$) |
| Average              | 22.6             | 15.0             | 7.5              | 33.2             |
| Median               | 18.5             | 10.9             | 7.4              | 29.3             |
| St.Dev*              | 12.5             | 11.7             | 2.7              | 17.3             |
| Max                  | 77.6             | 66.7             | 24.7             | 122.3            |
| Min                  | 6.7              | 2.4              | 0.3              | 6.5              |

*a St.Dev: Standard Deviation.

Fig. 1. (a) Map of Birmingham. The exact position of GB0851A station is marked with a red dot. (b) Birmingham Tyburn (GB0851A) air pollution station. The sampling site is highlighted with a red square.
Methodology

Computation of Backward Air Mass Trajectories

Hybrid Single-Particle Lagrangian Integrated Trajectory (HYSPLIT) model (http://www.arl.noaa.gov/HYSPLIT.php) of National Oceanic and Atmospheric Administration (NOAA) was used to produce 24-hour backward air mass trajectories based on NCEP/NCAR global reanalysis meteorological data. The air parcels reached the city of Birmingham at 12:00 UTC, during each one of the studied days, while the arrival altitude was set at 1000 m AGL (McGowan and Clark, 2008; Witkowska et al., 2015; Vadrevu et al., 2014). This upper boundary layer altitude was selected in order to implement the analysis at an intermediate height, between near-ground level (< 500 m AGL) and free troposphere (> 1500 m AGL) (Pawar et al., 2015), yet no significant differences are expected among backward trajectories starting within the atmospheric boundary layer (Adame et al., 2014; Lv et al., 2015).

Calculation of Trajectory Length

Each one of the computed trajectories is constituted by 25 hourly trajectory points (24 hours backward plus one arrival point). The coordinates (longitude/latitude) of the trajectory points are provided by HYSPLIT model. The length (L) of each atmospheric trajectory was calculated by an application of Haversine formula, as the sum of the 24 consecutive distances L_i (km) between the neighboring trajectory points along the atmospheric pathway (Markou and Kassomenos, 2010). Long range trajectories describe the movement of fast moving air parcels, implying increased wind speed, while short range trajectories are responsible for stagnant conditions and recirculation of polluted air (Zawadzka et al., 2013; Beddows et al., 2015).

Formulation of a Multiple Linear Regression Model

In various former publications, measured concentrations of gaseous pollutants and primarily those of NO_2 have been inserted as markers of combustion in linear regression models, in order to reconstitute the concentrations of airborne PM (Kuo et al., 2008; Liu and Harrison, 2011; Kassomenos et al., 2014; Wang et al., 2015). The main objective of this paper was to combine NO_2 levels, backward air mass trajectories and gridded satellite AOD data, as input (independent) variables in a Multiple Linear Regression Model (MLRM), purposing to improve significantly the estimation of PM_{10} concentrations. The proposed MLRM is described by Eq. (1):

\[ \text{PM}_{10}\text{d} = c_1[\text{NO}_2]\text{d} + c_2[L]\text{d} + c_3[AOD]\text{d–1} + c_4 \]  

(1)

where [PM_{10}\text{d}] is the estimated average daily PM_{10} concentration on day (d) and [NO_2]_d is the average daily concentration of NO_2 during day (d). [L]_d is the length of the 24-hr backward trajectory which reaches Birmingham on day (d), whereas [AOD]_{d–1} was calculated by averaging all the gridded AOD retrievals on day (d – 1) within 1° latitude and longitude from the trajectory pathway (Li et al., 2011).

Only days with available PM_{10}, NO_2 and AOD data were included in the study. In addition, aiming to enhance statistical stability, only [AOD]_{d–1} values calculated from at least 6 grid cells with available AOD measurements were used as model input. Thus, only 298 out of the 578 days of the studied period were finally studied.

Quantification of PM_{10} SOURCES and SINKS

The multiplication of the linear regression coefficients (c_1, c_2 and c_3) with the corresponding input variables (Eq. (1)), quantifies the contribution/deduction of particulates due to the influence of specific sources/wind dispersion respectively:

- c_1[NO_2]_d: PM_{10} contribution due to local combustion (c_1 > 0).
- c_2[L]_d: Removal of PM_{10} due to wind dispersion (c_2 < 0).
- c_3[AOD]_{d–1}: PM_{10} contribution due to long range transport (c_3 > 0)

Finally, the constant c_4 of the model is regarded as the contribution of local non-combustion PM_{10} emissions (c_4 > 0) (Harrison et al., 1997; Vardoulakis and Kassomenos, 2008; Liu and Harrison, 2011), attributed to dust resuspension, construction works, road friction, biogenic processes, etc.

The components (c_2[L]_d) and (c_3[AOD]_{d–1}) are both reflecting the influence of the incoming airflows on daily PM_{10} concentrations, however the opposite signs of the constants (c_2 < 0) and (c_3 > 0) indicate contrasting impacts. Hence, the overall effect of the prevailing atmospheric circulation is estimated by the combination of these two components. The final balance between the two terms is determined by the length of the air mass trajectories ([L]_d) and also by the volume of airborne aerosols ([AOD]_{d–1}) in close proximity to the atmospheric pathways.

Evaluation of the Model Outcome

Aiming to evaluate the overall performance of the proposed MLRM, a simple linear regression (least square method) and a Pearson Correlation Coefficient (PCC) analysis were performed among the observed (O\text{d}) daily PM_{10} concentrations and the predicted (P\text{d}) concentrations by the model (Li et al., 2011; Mishra et al., 2015). In addition, Mean Absolute Error (MAE, Eq. (2)) and Root Mean Square Error (RMSE, Eq. (3)) parameters were also calculated, in order to summarize the difference between the observed and modeled concentrations (Li et al., 2011; Dimitriou and Kassomenos, 2014b). In Eqs. (2) and (3), N is the total number of days and thus is equal to 298. In order to examine if the inclusion of atmospheric procedures (long range transport, wind dispersion) in the MLRM worked satisfactorily, PM_{10} concentrations were also estimated with another Simple Linear Regression Model (SLRM) which uses only NO_2 concentrations as a predictor variable. All the previously mentioned statistical tests were also applied for the SLRM.
\[ MAE = \frac{\sum |O_d - P_d|}{N} \]  \hspace{1cm} (2)

\[ RMSE = \sqrt{\frac{\sum (O_d - P_d)^2}{N}} \]  \hspace{1cm} (3)

RESULTS

Comparison of MLRM and SLRM

The results of MLRM and SLRM models are visualized in Figs. 2(a) and 2(b) respectively. According to PCC analysis, significant correlations at the 0.01 level were revealed among the observed PM\(_{10}\) concentrations and the predicted concentrations from both models. Thus, the impact of fuel combustion was clearly indicated from SLRM, yet the results of MLRM were substantially improved in comparison with the results of SLRM, as it was deduced from raised PCC values (Table 2). In addition, a stronger linear relation was identified between the monitored PM\(_{10}\) concentrations and the estimation of MLRM, according to R\(^2\) (R squared) values (Table 2, Fig. 3). Finally, reduced values of MAE and RMSE in the MLRM estimation (Table 2), also confirmed the better performance of the MLRM. Thus, the involvement of independent variables, reflecting atmospheric processes, clearly upgraded the prediction of PM\(_{10}\) levels through linear regression.

In general, modeled concentration data successfully follow the monitored PM\(_{10}\) levels however extreme values are underestimated (Fig. 2). Nevertheless, incidents of raised PM\(_{10}\) concentrations, attributed to long range transport impacts, were more effectively predicted by MLRM, thus the participation of long range transport variable in the MLRM was satisfactory. Two characteristic aerosol intrusions from transboundary sources are highlighted in Fig. 2.

Relative PM\(_{10}\) Contributions/Deductions

Atmospheric procedures (long range transport, wind dispersion) affecting PM\(_{10}\) levels in Birmingham, were quantified from MLRM, a possibility not previously available from SLRM. As it is shown from average daily

Fig. 2. (a) Observed PM\(_{10}\) concentrations (black color) and Predicted PM\(_{10}\) concentrations from MLRM (red color) (b) Observed PM\(_{10}\) concentrations (black color) and Predicted PM\(_{10}\) concentrations from SLRM (orange color). Characteristic long range transport episodes on 29/3/2011 and 24/5/2012 are highlighted.
Table 2. Statistical evaluation measures for MLRM and SLRM.

<table>
<thead>
<tr>
<th></th>
<th>MLRM</th>
<th>SLRM</th>
</tr>
</thead>
<tbody>
<tr>
<td>PCC</td>
<td>0.661*</td>
<td>0.477*</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.436</td>
<td>0.227</td>
</tr>
<tr>
<td>MAE (μgr m⁻³)</td>
<td>7.430</td>
<td>8.163</td>
</tr>
<tr>
<td>RMSE (μgr m⁻³)</td>
<td>9.876</td>
<td>11.561</td>
</tr>
</tbody>
</table>

*Values are statistically significant at the 0.01 level.

PM$_{10}$ contributions (Fig. 4(a)), anthropogenic combustion $(c_1[NO_2]_d)$ is the strongest source of PM$_{10}$ in the selected urban background area of Birmingham, followed by local non-combustion emissions $(c_4)$ and long range transport $(c_3[AOD]_{d-1})$. The mean daily deduction of aerosols, due to wind dispersion $(c_2[L]_d)$, was found equal to 9.3 μgr m⁻³ (Fig. 4(a)) but larger quantities of aerosols were removed from the atmosphere of Birmingham, during days characterized by the dominating effect of long range trajectories (i.e., increased wind speed) (Fig. 4(c)). Higher PM$_{10}$ contributions due to combustion were detected during winter and autumn (Fig. 4(b)), due to intensified vehicular and heating emissions. No clear seasonal trends were identified for PM$_{10}$ contribution/deduction attributed to long range transport/wind dispersion respectively (Figs. 4(c) and 4(d)).

**Extreme PM$_{10}$ Intrusions**

Multiple extreme intrusions of exogenous aerosols, enriching PM$_{10}$ levels in Birmingham Tyburn, have been indicated by the MLRM (Fig. 4(d)), corresponding to increased $[AOD]_{d-1}$ values. In accordance with previous publications, enhanced long range transport of PM$_{10}$ in Birmingham was mainly associated with the influence of

![Fig. 3. Scatter plots with simple linear regression lines among the observed PM$_{10}$ concentrations and the estimated PM$_{10}$ concentrations from (a) MLRM and (b) SLRM.](image-url)
Fig. 4. (a) Average daily PM$_{10}$ contributions/deduction from aerosol sources/wind dispersion respectively (b) Daily PM$_{10}$ contributions from local combustion (c) Daily PM$_{10}$ deductions from wind dispersion (d) Daily PM$_{10}$ contributions from long range transport.
air mass trajectories approaching from East–South East (E–SE) directions through Western and Central Europe (Borge et al., 2007; Vardoulakis and Kassomenos, 2008; Kavouras et al., 2013). Long range transport impacts were facilitated by the arrival of short range trajectories, due to low aerosol scattering (Beddows et al., 2015). In most cases, the prevalence of this type of atmospheric circulation provoked a radical increase of PM$_{2.5}$ concentrations (Kavouras et al., 2013), thus the incoming particulates are probably generated from urban and industrial combustion in continental Europe (Fotiadi et al., 2006; Minguillon et al., 2014). A characteristic invasion of PM$_{2.5}$ occurred in Birmingham on 29/3/2011 (Fig. 2, $c_3[AOD]_{\lambda=1} = 21.4\; \mu g\; m^{-3}$, PM$_{2.5} = 61.8\; \mu g\; m^{-3}$, PM$_{2.5}$/PM$_{10} = 87.3\%$), due to transportation of fine aerosols from Western France (Fig. 5(a)). However, during another aerosol intrusion ($c_3[AOD]_{\lambda=1} = 20.1\; \mu g\; m^{-3}$), which took place on 24/5/2012 (Fig. 2), PM$_{COARSE}$ concentrations were also elevated (PM$_{COARSE} = 18.2\; \mu g\; m^{-3}$, PM$_{2.5} = 27.3\; \mu g\; m^{-3}$, PM$_{2.5}$/PM$_{10} = 60.0\%$), probably due to the inflow of a mixture of exogenous aerosols (Fig. 5(b)), containing fine PM emitted from anthropogenic combustion and also maritime/crustal coarse particles (Fotiadi et al., 2006; Kocak et al., 2007).

CONCLUSIONS

This paper proposes a MLRM for the estimation of daily PM$_{10}$ concentrations in background urban areas. Apart from local combustion, the suggested method also intends to comprise the effect that transboundary aerosol sources and wind dispersion have, on particulate air pollution levels. However, anthropogenic combustion remains a primary cause

Fig. 5. Twenty-four hour backward atmospheric trajectories arriving in Birmingham on (a) 29/3/2011 and (b) 24/5/2012. Each trajectory is presented along with the previous day’s corresponding AOD (at 550 nm) observations from MODIS sensor. The city of Birmingham is marked with a black dot.
of increased PM$_{10}$ concentrations in urban areas, thus the relative effect of distinct aerosol sources and wind scattering was quantified by the MLRM’s variables. The MLRM was applied at a background urban site (Birmingham Tyburn) and the results were compared with the outcome of a SLRM which contained only one predictor variable expressing local combustion.

Several evaluation measures were used to examine the adequacy of the proposed method. All the implemented statistical tests signified the upgraded performance of the MLRM, in comparison with the SLRM, thus the influence of the prevailing atmospheric circulation was sufficiently expressed by the model. Moreover, the inclusion of the long range transport variable in the MLRM indicated various intense intrusions of exogenous particles in Birmingham, mostly associated with the advection of fine (PM$_{2.5}$) particles from continental Europe. The estimation of PM$_{10}$ concentrations by the MLRM, during those extreme long range transport events, was particularly improved versus the SLRM. However, as it was deduced from the MLRM, local combustion (traffic, heating) is an important PM$_{10}$ contributor in Birmingham Tyburn, thus the results of the SLRM were also statistically significant.

Conclusively, the involvement of the selected predictive factors in a MLRM, successfully upgraded the estimation of PM$_{10}$ concentrations. Thus, key elements of this model can be used in order to organize PM abatement strategies and prevent human exposure. Nevertheless, the proposed technique was only applied in a background urban site of Birmingham, a city in which the impact of aerosol transportation has been previously reported. In addition, the need for ground based air pollution data, in conjunction with sufficient satellite AOD observations, decreased the number of suitable days for the implementation of the MLRM. Hence, more work is required, purposing to investigate the effectiveness of this method with larger datasets in other urban areas. Finally, this approach can be further developed in the future by the addendum of more independent variables which will further enhance the model’s performance.

ACKNOWLEDGEMENTS

The author would like to thank the EU Air Quality Database, for the concession of air pollution data, and also the MODIS science data support team for providing data via the Giovanni website (http://giovanni.gsfc.nasa.gov). The author gratefully acknowledges the NOAA Air Resources Laboratory (ARL) for the provision of the HYSPLIT transport and dispersion model used in this publication. Information for the air pollution station GB0851A (Birmingham Tyburn) were found on the website of the Department for Environment Food and Rural Affairs (http://uk-air.defra.gov.uk). Finally I would like to thank professor Pavlos Kassomenos for his guidance and support.

REFERENCES


Received for review, May 27, 2016
Revised, August 9, 2016
Accepted, August 9, 2016