JOURNAL OF ENVIRONMENTAL SCIENCES 150 (2025) 676–691

Available online at www.sciencedirect.com

ScienceDirect

www.elsevier.com/locate/jes

JES JOURNAL OF ENVIRONMENTAL SCIENCES

www.jesc.ac.cn

A methodological framework for estimating ambient $PM_{2.5}$ particulate matter concentrations in the UK

David Galán-Madruga[1,**], *Parya Broomandi*[2,3,**], *Alfrendo Satyanaga*[2], *Ali Jahanbakhshi*[4], *Mehdi Bagheri*[3], *Aram Fathian*[5,6,7], *Rasoul Sarvestan*[8], *J. Cárdenas-Escudero*[9,10], *J.O. Cáceres*[9], *Prashant Kumar*[11,12], *Jong Ryeol Kim*[2,*]

[1] *Department of Atmospheric Pollution, National Centre for Environment Health, Health Institute Carlos III. Ctra. Majadahonda a Pozuelo km 2.2, Madrid 28220, Spain*
[2] *Department of Civil and Environmental Engineering, School of Engineering and Digital Sciences, Nazarbayev University, Kabanbay Batyr Ave. 53, Astana 010000, Kazakhstan*
[3] *Department of Electrical and Computer Engineering, School of Engineering and Digital Sciences, Nazarbayev University, Kabanbay Batyr Ave. 53, Astana 010000, Kazakhstan*
[4] *School of Architecture, Building and Civil Engineering, Loughborough University, Leicestershire LE11 3TU, United Kingdom*
[5] *Neotectonics and Natural Hazards Institute, RWTH Aachen University, Aachen 52056, Germany*
[6] *UNESCO Chair on Coastal Geo-Hazard Analysis, Research Institute for Earth Sciences, Tehran 13185-1494, Iran*
[7] *Water, Sediment, Hazards, and Earth-surface Dynamics (waterSHED) Lab, Department of Geoscience, University of Calgary, Calgary Alberta T2N 1N4, Canada*
[8] *Department of Climatology, Faculty of Geography and Environmental Sciences, Hakim Sabzevari University, Sabzevar 9617976487, Iran*
[9] *Laser Chemistry Research Group, Department of Analytical Chemistry, Faculty of Chemistry, Complutense University of Madrid, Plaza de Ciencias 1, Madrid 28040, Spain*
[10] *Analytical Chemistry Department, FCNET, University of Panama, University City, University Mail, Panama 4, Panama City 3366, Panama*
[11] *Global Centre for Clean Air Research (GCARE), School of Sustainability, Civil and Environmental Engineering, Faculty of Engineering and Physical Sciences, University of Surrey, Guildford GU2 7XH, United Kingdom*
[12] *Institute for Sustainability, University of Surrey, Guildford GU2 7XH, United Kingdom*

ARTICLE INFO

Article history:
Received 24 April 2023
Revised 17 November 2023
Accepted 19 November 2023
Available online 27 November 2023

ABSTRACT

Scientific evidence sustains $PM_{2.5}$ particles' inhalation may generate harmful impacts on human beings' health; therefore, their monitoring in ambient air is of paramount relevance in terms of public health. Due to the limited number of fixed stations within the air quality monitoring networks, development of methodological frameworks to model ambient air $PM_{2.5}$ particles is primordial to providing additional information on $PM_{2.5}$ exposure and its trends. In this sense, this work aims to offer a global easily-applicable tool to estimate ambient air $PM_{2.5}$ as a function of meteorological conditions using a multivariate analysis. Daily $PM_{2.5}$ data measured by 84 fixed monitoring stations and meteorological data from ERA5 (ECMWF Reanalysis v5) reanalysis daily based data between 2000 and 2021 across the United

** These two authors contributed equally to this work.
* Corresponding author.
E-mail: jong.kim@nu.edu.kz (J.R. Kim).

https://doi.org/10.1016/j.jes.2023.11.019
1001-0742/© 2024 The Research Center for Eco-Environmental Sciences, Chinese Academy of Sciences. Published by Elsevier B.V.

Keywords:
Air quality
$PM_{2.5}$ particles
Meteorological variables
Prediction model
Long-term trend

Kingdom were attended to develop the suggested approach. Data from January 2017 to December 2020 were employed to build a mathematical expression that related the dependent variable ($PM_{2.5}$) to predictor ones (sea-level pressure, planetary boundary layer height, temperature, precipitation, wind direction and speed), while 2021 data tested the model. Evaluation indicators evidenced a good performance of model (maximum values of *RMSE, MAE* and *MAPE*: 1.80 μg/m^3, 3.24 μg/m^3, and 20.63%, respectively), compiling the current legislation's requirements for modelling ambient air $PM_{2.5}$ concentrations. A retrospective analysis of meteorological features allowed estimating ambient air $PM_{2.5}$ concentrations from 2000 to 2021. The highest $PM_{2.5}$ concentrations relapsed in the Mid- and Southlands, while Northlands sustained the lowest concentrations.

© 2024 The Research Center for Eco-Environmental Sciences, Chinese Academy of Sciences. Published by Elsevier B.V.

Introduction

Airborne particles influence the global climate and ecosystem processes according to their physicochemical characterizations. By addressing the negative health impacts of being exposed to airborne particles (long-term and/or short-term) in many studies, the importance of more efficient abatement measures is highlighted (Annesi-Maesano et al., 2007; Chen et al., 2020; Crumeyrolle et al., 2014; Duncan et al., 2014; Guevara-Luna et al., 2018; Jiang and Christakos, 2018; Krishna et al., 2019; Nabavi et al., 2018; Wongwatcharapaiboon, 2020; Xu and Zhang, 2020; Zeydan and Wang, 2019).

In the EU28 countries, based on a report of the European Environment Agency, around 10%–14% of the urban people were exposed to $PM_{2.5}$ concentrations exceeding the EU annual-mean $PM_{2.5}$ reference value (25 μg/m^3), while 91%–93% were exposed to $PM_{2.5}$ levels exceeding the WHO annual-mean $PM_{2.5}$ (10 μg/m^3) from 2010 to 2012 (EEA, 2014; Gehrig and Buchmann, 2003; Gehring et al., 2015). The results of the Global Burden of Disease project estimated about 76 million disability-adjusted life years and 3.2 million premature casualties caused by being exposed to $PM_{2.5}$ concentrations prevailing in 2005 (Lim et al., 2012; Vieno et al., 2016). In the United Kingdom, long-term exposure to PM mainly from anthropogenic sources causes around 29,000 causalities in a year (COMEAP, 2010; Gowers et al., 2014).

Historical $PM_{2.5}$ datasets with high spatial-temporal resolutions play a vital role in evaluating the long-term impacts of $PM_{2.5}$ on human health, the atmospheric environment, and climate change. Generally, $PM_{2.5}$ monitoring can be divided into two groups: (a) ground-based monitoring stations, and (b) satellite remote sensing image estimation (Lin et al., 2015; Xu and Zhang, 2020). The obtained air quality data from ground monitoring stations are suffering from many limitations including limitations in number, unequal distribution, and having different measure frequency ranges (Chu et al., 2016; Krishna et al., 2019; Xiao et al., 2017). Previously aforementioned obstacles can impact the demographical and geographical range of research, causing an information bias and reducing confidence in the results of exposure-response research (Liu et al., 2007). Moreover, there is a complexity in the spatial and temporal variation of ambient $PM_{2.5}$ and continuous monitoring of $PM_{2.5}$ levels is not present in many regions and countries (Zhang and Cao, 2015). The hourly $PM_{2.5}$ monitoring network in the UK (Department of Environment Food and Rural Affairs) was not established until 2000, only covering a very limited number of air quality stations between 2000 and 2007 (https://uk-air.defra.gov.uk/data/openair) (DEFRA, 2015). Therefore, retrieving historical $PM_{2.5}$ datasets due to the limited and sparse $PM_{2.5}$ observations, is a research challenge.

To overcome this challenge, satellite-retrieved aerosol optical depth (AOD) with relatively long observation periods (∼ 20 years) and broad spatial coverage has been increasingly used to estimate large-scale $PM_{2.5}$ concentrations for the last twenty years (Geng et al., 2020, 2018; Huang et al., 2018; Shin et al., 2020; Xiao et al., 2018, 2017; Zhong et al., 2021). It helps us to obtain continuous and large-scale atmospheric aerosol data. These data can then be converted into the $PM_{2.5}$ concentration and distribution at the earth's surface. However, high-precision regression models need to be established between ground-based $PM_{2.5}$ concentration and satellite-based atmospheric particle data (Huang et al., 2018; Ma et al., 2014; Zhao et al., 2018). Yazdi et al. (2020) created a prediction model for daily $PM_{2.5}$ concentrations using AOD (Aerosol Optical Depth) from the satellite data in the Greater London area using an ensemble machine learning by adding meteorological parameters, and land use. Another study developed a multi-stage satellite-based Machine Learning model to estimate $PM_{2.5}$ concentrations across the UK in the period 2008-2018. To have a precise estimation, they added meteorological parameters, land use information, and site-specific and emission inventory to the statistical analysis (Schneider dos Santos et al., 2020).

However, some inherent limitations are associated with satellite-based AOD which are hard to overcome. For instance, availability is significantly affected by the large proportion of non-random missing AOD due to cloud cover, which generates biases (Bi et al., 2019; Xiao et al., 2017). Previous studies have also estimated that around 70%-90% of total AOD retrievals from MODIS (Moderate Resolution Imaging Spectroradiometer) data are missing (Bi et al., 2019; Chen et al., 2019). Besides, the sampling frequency of satellite-based AOD is typically limited to a maximum of twice a day. This limitation makes hourly $PM_{2.5}$ impossible and causes an underrepresentation of the average daily AOD. In model accuracy, outside the training period, the predictive accuracy of samples is significantly lower than the validation accuracy, indicating a relatively weak predictive capability of these models (Wei et al., 2019; Xiao et al., 2018).

While meteorological data have a unique advantage compared with satellite-based AOD in retrieving historical $PM_{2.5}$ datasets. These data can be easily traced back to the 1950s with much more extended temporal resolution than satellite-based AOD observations. In the case study, they can also be acquired from ERA5 (ECMWF Reanalysis v5) reanalysis daily based data. Surface meteorological observations are not affected by cloud cover and can continuously be recorded, overcoming the shortcomings of satellite-based AOD data with low sample frequency and massive missing values.

Despite $PM_{2.5}$ stations which are mainly installed in cities, meteorological stations are distributed more densely and evenly, which provides a ground for retrieving historical $PM_{2.5}$ datasets at high spatial-temporal resolutions using surface meteorological variables. As previously stated, when there is a lack of sufficient meteorological stations, the necessary data can be obtained from ERA5 (ECMWF Reanalysis v5) reanalysis, which provides daily data.

The aim of this work is to build a globally applicable tool that provides information on ambient air $PM_{2.5}$ concentrations, given that $PM_{2.5}$ is of paramount relevance due to its potentially harmful effect on human health (Heal et al., 2012). Therefore, their assessment is necessary in European Member States. Based on that, the aim of this work is to evaluate whether modelling $PM_{2.5}$ levels from surface meteorological observations can reliably estimate ambient air levels of $PM_{2.5}$ particles compared to actual levels.

As a case study, the United Kingdom was considered an investigating domain for the proposed objective. Mathematical equations were obtained as a function of North-, Mid-, and Southlands. As a second milestone, a retrospective analysis on $PM_{2.5}$ concentrations in the United Kingdom from 2000 to 2021 using meteorological data was performed because historical long-term $PM_{2.5}$ records are essential to (i) know and understand its long-time trend, (ii) characterize behaviors, (iii) perform epidemiological studies, (iv) assess its human health impact, (v) evaluate mitigation plans or/and air quality control strategies, among other topics.

Therefore, to achieve our goal, we proposed an optimized mathematical model for estimating seasonal average $PM_{2.5}$ concentrations using surface meteorological observations, which will enrich the $PM_{2.5}$ data availability and contribute to better evaluating the potential risk associated with fine particles ($PM_{2.5}$) human exposure. As a practical application, a long-time trend study is conducted using the suggested approach to estimate seasonal $PM_{2.5}$ particle concentrations. As a consequence, long-time spatiotemporal variations of ambient $PM_{2.5}$ particles are discussed, and a densely gridded information network is established within the target study domain, which is fundamental to protecting human health.

1. Materials and methods

1.1. *Original database concerning ground-level measurements*

The United Kingdom is taken as a domain area. It consists of four countries (Scotland, England, Wales and Northern Ireland) with a total population of around 67.1 million in 2018 (ONS, 2020). In terms of climate classification, the United Kingdom is classified as fully humid with mostly warm summers (Kottek et al., 2006) and occupies 243,610 km^2 (Fig. 1).

To develop the present investigation, air quality and meteorological conditions data are required. In this sense, as a specific support, daily air quality data for $PM_{2.5}$ particles measured by fixed monitoring stations between 2000 and 2021 across the United Kingdom were acquired from the Kingdom-air Defra dataset website (https://uk-air.defra.gov.uk/data/openair) (DEFRA, 2015). Appendix A Table S1 contains relevant information on fixed stations involved in this study in terms of the type of location (urban, suburban and rural), main pollution source (industrial, traffic and background), geographical location and elevation above sea level. A total of 84 fixed stations for measuring $PM_{2.5}$ particles were involved.

The daily meteorological data involving mean sea-level pressure (MSLP), planetary boundary layer height (PBLH), temperature (T), precipitation (P), wind direction (WD) and speed (WS) were attended (covered period: 2000-2021). They were acquired from ERA5 (ECMWF Reanalysis v5) reanalysis daily based data. ERA5 reanalysis data is produced by Copernicus Climate Change Service (C3S) at European Centre for Medium-Range Weather Forecasts (ECMWF) (ERA5, 2020). Appendix A Table S2 presents the summary of meteorological data.

The following criteria were considered regarding the temporal frame and study domain: (i) Within the researched period, each studied year was split into four parts coinciding with the meteorological seasons (Winter, from January 1 to March 30, Spring, from April 1 to June 30, Summer, from July 1 to September 30, and Autumn, from October 1 to December 31); and (ii) Within the studied surface, three sub-area were considered (North-, Mid-, and South-lands).

Moreover, all the spatial representation and interpolation techniques were carried out through the QGIS (Version 3.22.3) software. The IDW (Inverse Distance Weighting) interpolation algorithm was applied with a spatial resolution of 0.1° by 0.1° for the output raster.

1.2. *Modeling ambient air $PM_{2.5}$ particle levels from surface meteorological observations*

The key objective of this analysis was to reach a mathematical expression that lays down the relationship between seasonal average values of ambient air $PM_{2.5}$ concentrations and surface meteorological variables, given that these play a relevant function in the dispersion dynamic of air pollutants. Previous investigation studies have already reported the paramount of meteorological variables on atmospheric pollution forecasting (Cortina-Januchs et al., 2015).

To achieve this scope, a $PM_{2.5}$ and meteorological data subset from the original database was employed (temporary frame: from January 2017 to December 2020). To incorporate all study variables within a multivariate analysis, a multiple linear regression test (MLR) was conducted, in concordance with other scientific studies (Kumar and Goyal, 2011). In this context, $PM_{2.5}$ was considered the dependent variable and meteorological features (MSLP, PBLH, T, P, WD and WS) as the predictor variables. Then, a mathematical model was rigged up to estimate $PM_{2.5}$ concentrations from the predictor variables.

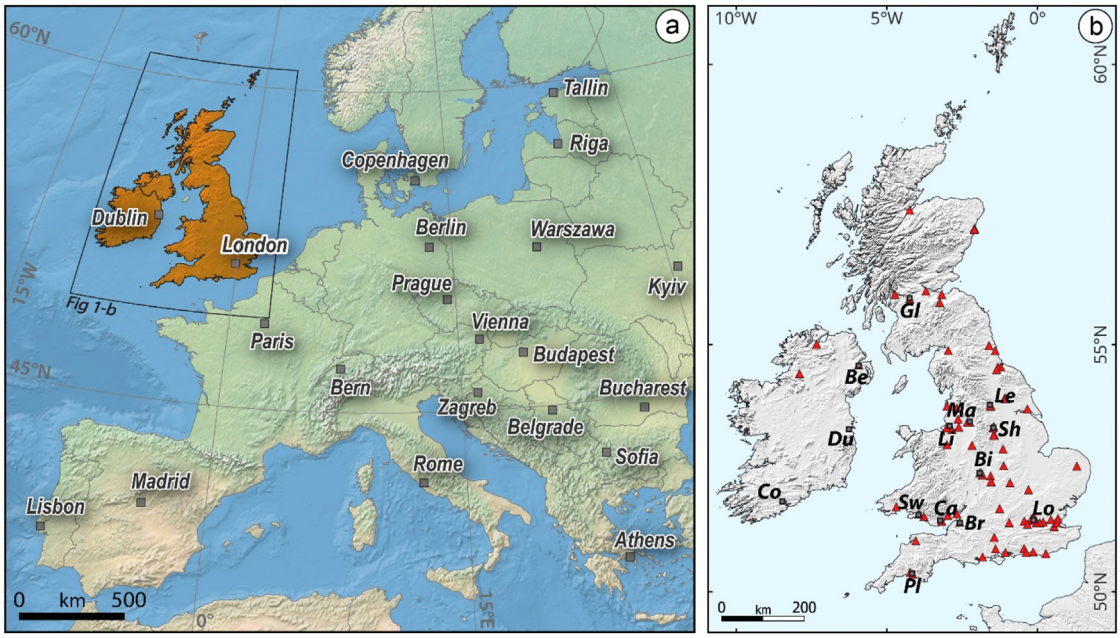

Fig. 1 – The investigated domain in the current study from 2000 to 2021.

To further enhance the reliability of the constructed model, the approach reported by Eurachem (2012) was applied. Shortly, the pragmatic relationship between a reference and candidate procedure is given by bias determination. In this study, the quantification of bias associated with the model was calculated as the difference between the seasonal average of the estimated (using the proposed mathematical model) and actual $PM_{2.5}$ concentrations. Then, every model equation (seasonal and territorial) was adjusted with the average value of the bias.

The multivariate analysis approach developed to estimate ambient air $PM_{2.5}$ concentrations were designed to account for seasonal and territory criteria. The following mathematical formula establishes the relationship between the dependent and predictor variables.

$$PM_{2.5\,E} = Cte + ax_1 + bx_2 + cx_3 + dx_4 + ex_5 + fx_6 + \varepsilon \quad (1)$$

where, Cte is a constant value (it corresponds to the intercept); a, b, c, d, e and f are mixed effect regression (they correspond to the scopes); from x_1 to x_6 are the predictor variables (MSLP, PBLH, T, P, WD and WS, respectively), and ε is the bias associated with the proposed estimate model.

1.3. *Validation of the suggested mathematical model to estimate ambient air $PM_{2.5}$ levels*

Mathematical formulas to estimate seasonal $PM_{2.5}$ concentrations in ambient air were obtained by executing the sequence cited in the previous section. To evaluate the validity of the suggested equations, seasonal $PM_{2.5}$ particle levels were estimated during the 2021 year in the function of the division of the target territory (North-, Mid-, and South-lands).

To value the confidence of the recommended approach, a simple linear regression technique was executed between the estimated (dependent variable) and actual concentration (independent variable) of ambient air $PM_{2.5}$ particles during 2021, taking into account the following evaluation requirements: (i) measure of the correlation ratio between estimated and actual $PM_{2.5}$ concentrations, (ii) statistical significance of the simple linear regression expression with one-way ANOVA, (iii) statistical significance of the regression coefficient of the independent variable, and (iv) statistical significance between estimated vs actual samples, as evaluated by a t-test (Yang et al., 2018). Note that the modeling period covered 1 January 2017 to 31 December 2020.

Moreover, to test this hypothesis, the proposed approach was applied to another air quality network located in another geographical zone with different emission air pollutant focuses, meteorological conditions and topographical parameters. In concrete, two fixed monitoring stations belonging to Madrid City's air quality network were considered. These stations are catalogued as suburban stations (Casa de Campo, Code: 28079024, Latitude: 40°25′09.68′′N, Longitude: 3°44′50.45′′W) and urban traffic stations (Plaza Elíptica, Code: 28079056, Latitude: 40°23′05.87′′ N, Longitude: 3°43′07.42′′ W). Therefore, these stations present distinct polluting profiles. Current $PM_{2.5}$ concentrations and surface meteorological values for each station were downloaded from Madrid City's open data portal (http://www.mambiente.munimadrid.es). Seasonal average meteorological variables values measured by the stations sited in Casa de Campo and Plaza Elíptica were used to apply the global meteorological formulae previously indicated.

To evaluate the model, three indicators widely used in diverse disciplines, as atmospheric sciences (Karunasingha, 2022), were regarded to assess the proposed model performance, named: root mean square error (*RMSE*), mean prediction error (*MAE*) and mean absolute percentage error (*MAPE*), which were estimated using the

following equations, according to Dai et al. (2022):

$$RMSE = \sqrt{\frac{\sum (PM_{2.5\ E} - PM_{2.5\ C})^2}{N}} \quad (2)$$

$$MAE = \frac{\sum (PM_{2.5\ E} - PM_{2.5\ C})^2}{N} \quad (3)$$

$$MAPE = \frac{\sum /PM_{2.5\ E} - PM_{2.5\ C}/}{N} X \frac{100\%}{PM_{2.5\ C}} \quad (4)$$

where, $PM_{2.5\ E}$ (μg/m^3) and $PM_{2.5\ C}$ (μg/m^3) correspond to the estimated and current $PM_{2.5}$ concentration, respectively, and N equals to the number of samples.

1.4. *Retrospective analysis of the time series data streams corresponding to the predictor variables*

This analysis was carried out to address the seasonal occurrence of $PM_{2.5}$ particles in the target domain from 2000 to 2021 as a practical application of the reported estimate approach. For that, a retrospective predictor variables analysis is previously required. In this sense, the application of simple linear regression-based analysis is often employed to estimate unknown values concerning a specific target variable. The time series data streams are studied to understand the subjacent behavior of the dataset within a concrete time interval, building a fitting line between a dependent vs independent variable (Barman and Choudhury, 2020). Therefore, this approach can quantitatively estimate either preterit or future values concerning a time series based on its historical pattern.

The approach proposed by Galán-Madruga (2022) was adopted to run a retrospective study of the time series of each predictor variable. To achieve this goal, seasonal average values of predictor variables were considered. Given that the previously cited scientific work consecutively listed the points needed for conducting a time series analysis, now it is cited as the only difference regarding the original approach. In this regard, Galán-Madruga (2022) used to perform a prospective investigation, whereas we used a preterit study. For this reason, within the first point of the original approach, the time variable was converted to a numeric value, following an increasing order, unlike the present study, where a decreasing order is implemented (retrospective analysis). Therefore, the most elevated numeric value corresponds to the last seasonal period and year where predictor variables data are monitored for each meteorological station following in decreasing order up to the 2000 year. The not-monitored years were estimated and compared to those monitored by target stations.

1.5. *$PM_{2.5}$ Time Series Analysis*

To analyze the significance of trends in long-term changes in $PM_{2.5}$ data, Mann–Kendall test ("(wq)" package in R) was applied, using the "seaKen" functions (Jassby et al., 2016), for the seasonal Mann–Kendall trend tests (Kendall, 1949; Mann, 1945). The Mann–Kendall trend test is a specific nonparametric technique often applied to indicate the possible trend for time series data. Moreover, it can represent the seasonal patterns within the data. The Sen's Slope estimator was also applied, from the same package (wq), to quantify the magnitude of changes in $PM_{2.5}$ concentrations. Considering the autocorrelation limitation, the test was performed using monthly averages. Because they are far apart in the time step, autocorrelation is unlikely to be crucial (Silva Junior et al., 2018).

2. Results and Discussion

2.1. *Modeling ambient air $PM_{2.5}$ concentrations from surface meteorological observations*

A descriptive analysis is carried out to provide relevant information on the status of predictor variables over the modelling period (2017-2020). Seasonal average values of the predictor variables (MSLP, PBLH, T, P, WD and WS in this order) were 1013.97 hPa (1003.62 and 1019.80 hPa minimum and maximum, respectively), 686.75 m (416.86-945.24 m), 10.59°C (2.43-17.94°C), 2.54 mm/day (0.73-7.80 mm/day), 136.84 degrees (75.79-215.95 degrees) and 4.32 m/sec (2.73-8.82 m/sec). A more detailed analysis evidenced profiles reasonably similar to MSLP, PBLH and WD in the function of the North-, Mid-, and South-lands (relative maximum differences lower than 20%) while T, P and WS sustained 43.87%, 25.45% and 32.99% maximum variations, respectively (Appendix A Fig. S1).

Appendix A Table S3 shows a study of linearity between $PM_{2.5}$ concentrations and meteorological parameters (2017-2020). Note that PBLH, P and T have a negative correlation concerning the dependent variable. Appendix A Table S4 displays the outcomes as a function of seasonal period and territorial division, by applying the MLR technique over the 2017-2020 period. Note that the suggested multivariate analysis provided a good capability for predicting $PM_{2.5}$ levels in ambient air sustained by Pearson's coefficients ranging between 61.3% and 91.1%, with the highest predictive efficiency in the winter period (higher than 70.0%) followed by the rest of seasonal periods (60% < predictive capability < 70%), except for the springtime in the Midlands (83.6%). To further knowledge of the Pearson's coefficients outcomes obtained by the MLR technique, the interpretation proposed by Dancey and Reidy (2007) was adopted. They employed the following intervals for ranking the Pearson's coefficients values: 0 (it corresponds to an association degree between the correlated variables of zero), ± 0.1 to ± 0.3 (weak correlation), ± 0.4 to ± 0.6 (moderate correlation), ± 0.7 to ± 0.9 (strong correlation) and ± 1.0 (perfect correlation). To lay down a comparison between Pearson's coefficients reported in Table 1 and the previously listed classification, those attained in this study are rounded to one decimal. Based on that, moderate and strong associations accounted for 42% and 58%, respectively. Another aspect that supports the proper development of an estimation model is the independence among residues. In this case, this requirement is guaranteed, given that the Durbin–Watson statistic fell into the value interval between 1.5 and 2.5 (Appendix A Table S4), which is in line with the performance of other estimating models reported by other investigation studies (Galán-Madruga and García Cambero, 2022). It is worth highlighting that they did not execute outlier tests before the data treatment (Galán-Madruga and García Cambero, 2022). Therefore, all the data from the original database was included in the

Table 1 – General multiple linear regression equations for estimating ambient air seasonal average $PM_{2.5}$ levels from target predictor variables.

Target surface	Seasonal period	Mathematical equation
Northlands	Winter	$PM_{2.5} = -474.321 + (0.477 \times MSLP) - (0.003 \times PBLH) - (0.540 \times P) - (0.366 \times T) + (0.033 \times WD) + (0.584 \times WS)$
	Spring	$PM_{2.5} = 1484.224 - (1.469 \times MSLP) - (0.004 \times PBLH) - (0.790 \times P) + (1.139 \times T) + (0.037 \times WD) + (0.360 \times WS)$
	Summer	$PM_{2.5} = 267.527 - (0.266 \times MSLP) - (0.010 \times PBLH) - (0.520 \times P) + (1.322 \times T) - (0.014 \times WD) - (0.339 \times WS)$
	Autumn	$PM_{2.5} = -103.559 + (0.120 \times MSLP) - (0.037 \times PBLH) - (0.485 \times P) + (1.518 \times T) + (0.002 \times WD) + (0.675 \times WS)$
Midlands	Winter	$PM_{2.5} = 505.298 - (0.474 \times MSLP) - (0.043 \times PBLH) - (1.450 \times P) + (1.363 \times T) + (0.022 \times WD) + (2.461 \times WS)$
	Spring	$PM_{2.5} = 1912.564 - (1.886 \times MSLP) - (0.005 \times PBLH) + (0.297 \times P) + (1.437 \times T) - (0.008 \times WD) + (0.099 \times WS)$
	Summer	$PM_{2.5} = -110.221 + (0.097 \times MSLP) + (0.055 \times PBLH) + (1.389 \times P) - (1.486 \times T) + (0.037 \times WD) - (1.775 \times WS)$
	Autumn	$PM_{2.5} = 24.379 + (0.003 \times MSLP) - (0.006 \times PBLH) - (1.931 \times P) + (1.485 \times T) + (0.013 \times WD) + (0.373 \times WS)$
Southlands	Winter	$PM_{2.5} = 712.109 - (0.673 \times MSLP) - (0.037 \times PBLH) - (0.493 \times P) + (2.248 \times T) - (0.012 \times WD) - (0.407 \times WS)$
	Spring	$PM_{2.5} = 663.396 - (0.649 \times MSLP) + (0.007 \times PBLH) + (0.325 \times P) - (0.078 \times T) + (0.010 \times WD) + (0.748 \times WS)$
	Summer	$PM_{2.5} = 307.046 - (0.294 \times MSLP) - (0.009 \times PBLH) + (0.458 \times P) + (0.923 \times T) - (0.039 \times WD) - (1.414W \times S)$
	Autumn	$PM_{2.5} = 11.648 - (0.002 \times MSLP) - (0.016 \times PBLH) - (0.220 \times P) + (1.689 \times T) - (0.002 \times WD) - (0.663 \times WS)$

Note: MSLP (mean sea-level pressure) expressed in hPa, PBLH (planetary boundary layer height) in m, P (precipitation) in mm/day, T (temperature) in °C, WD (wind direction) in degrees and WS (wind speed) in m/sec.

modeling $PM_{2.5}$ process using the target surface meteorological features.

A primordial activity in modeling processes leads to the evaluation of bias associated with the proposed estimating approach. In this case, reference versus the proposed model bias was assessed as a function of the target study surface and seasonal period. Mean bias values (in μg/m^3) of -0.005 (relative value: -4.729%), 0.727 (4.267%), 0.172 (-1.331%) and 0.444 (-1.251%) for winter, spring, summer and autumn, respectively, in the Northlands, were obtained according to bias calculus description in Section 2.3. Similarly, a bias of -0.596 (6.544%), 0.622 (5.212%), 0.732 (8.455%) and -0.360 (-4.954%) in Midlands and 0.129 (1.442%), -0.655 (-8.869%), -0.122 (-6.989%) and -0.322 (-7.349%) in Southlands were averaged. These mean bias values were included in the mathematical formulas to enhance the predictive capacity of the suggested estimation model. In this sense, Table 1 shows the general multiple linear regression equations for estimating ambient air seasonal average $PM_{2.5}$ levels using surface meteorological features as predictor variables.

Given that the dispersion of pollutants in the air matrix is highly conditioned by meteorological characteristics (Ramsey et al., 2014), the existing relationship between the target air pollutant and each involved meteorological condition was assessed over the modeling period (2017-2020) (Appendix A Table S5). To favor the compression of the reported outcomes, these were grouped according to Dancey and Reidy's coefficient of correlation classification. Broadly, it can be concluded that the results sustain a weak association between current $PM_{2.5}$ levels and meteorological variables (with ∼ 67% of the total of the studied associations, Appendix A Fig. S2A). More detailed analysis as a function of the individual meteorological variables pointed to P, WD and WS as the predictor variables with the highest contribution level to this type of association. In contrast, MSLP exhibited a strong connection (25%) concerning current $PM_{2.5}$ concentrations whereas PBLH (33%) and T (67%) showed a moderate correlation (Appendix A Fig. S2B).

In executing the same analysis exercise considering the target surface, the Northlands were revealed as the study surface where the meteorological conditions greatly influence current $PM_{2.5}$ concentrations in ambient air (with particular emphasis on MSLP), followed by the Midlands (particularly WS) and Southlands (mainly MSLP, PBLH and T), respectively (Fig. 2).

Due to the potentially damaging impact of $PM_{2.5}$ exposure on human health, the development of predicting models that help evaluate human exposure to fine particles has been a great scientific concern. In this sense, scientific literature collects numerous research works focused on forecasting this target pollutant in ambient air. In this sense, different analysis techniques and predictor variables have been considered for estimating $PM_{2.5}$ concentrations.

Kumar et al. (2020) built a predicting $PM_{2.5}$ model applying time series analysis and regression techniques, being the study period from 1 January 2018 to 30 November 2019 using 80% data for learning (model training) and 20% for testing the outcomes. They downloaded seven meteorological parameters, from the monitoring station set up in the R K Puram area, Delhi (India), as predictor variables. The estimated model showed a predictive 92.64% capacity. Other distinguished research groups based their prediction of $PM_{2.5}$ models on more complex techniques such as machine learning approaches. For example, Masood and Ahmad (2020) used an artificial neural network for predicting $PM_{2.5}$ concentrations by employing six meteorological features and six air pollutants as independent variables in Delhi, India (May 2016-May 2018: 80% of data was used for model training and remaining 20% for validating process). Pearson's coefficients of correlation of 0.856 for the model training and 0.730 for the testing phase (artificial neural networks) sustained a good estimate capacity. Doreswamy et al. (2020) elaborated a $PM_{2.5}$ model for the 2012-2017 period using particular and chronological data from Pingtung station, Newport City in Taiwan, China. In their study, chronological data pertains to information that undergoes rapid changes within a short period of time. Particular parameters are factors that vary both spatially and temporally, resulting in fluctuations in the concentration of air pollution. They applied algorithms such as linear regression, random forest regressor, gradient boosting regression and deci-

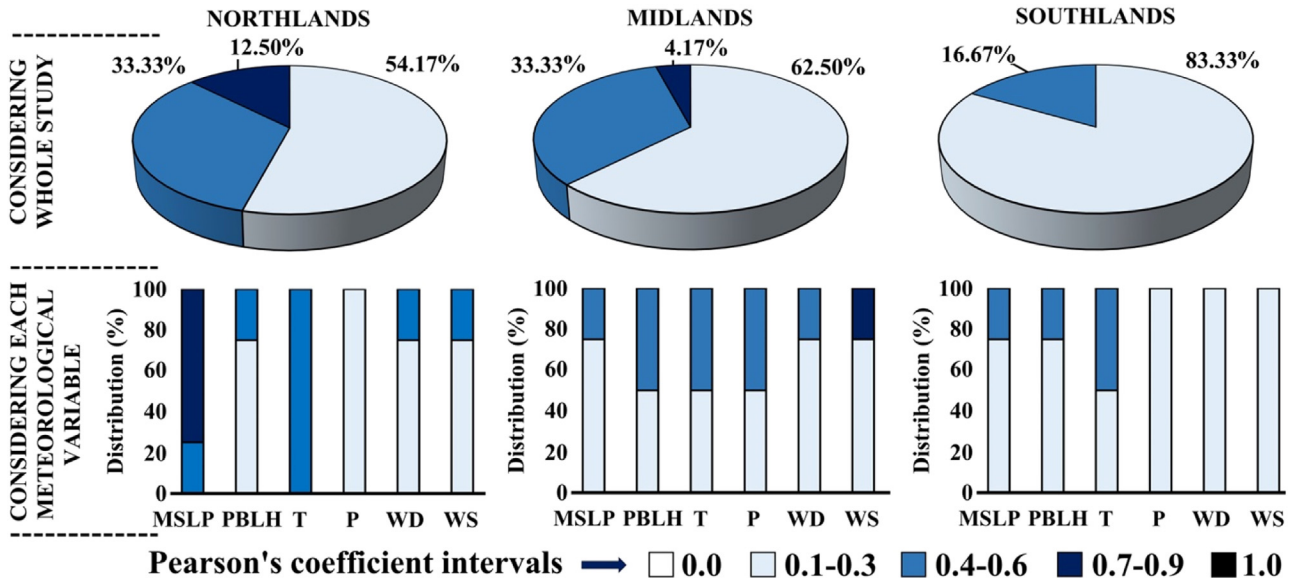

Fig. 2 – Distribution of Pearson's coefficients when correlating current $PM_{2.5}$ levels versus meteorological conditions (2017-2021) as a function of target study surface.

sion tree regression, obtaining coefficients of determination's values of 0.793, 0.987, 0.998 and 0.999, respectively, for the model training and 0.720, 0.870, 0.889 and 0.799, respectively for the testing process. Peng et al. (2022) used $PM_{2.5}$ concentrations recorded between 1 January and 31 December 2021, at two monitoring sites in Changsha, China, Hunan Province, to estimate $PM_{2.5}$ levels by applying the XGBoost model. They reached a coefficient of determination of up to 0.856.

The remarkable previously cited research works explored mean absolute error (it establishes the horizontal separation between two continuous values) and root mean squared error (it relates the standard deviation of the predicted concentrations from real levels) as metrics for assessing the performance of the proposed models, expressing the results in nominal scale (μg/m^3) against a relative scale (%). This fact may lead to interpretation errors, given that the nominal scale does not provide a clear overview of the real difference between predicted and current target pollutant levels. As an example, an original dataset comprised of low real $PM_{2.5}$ concentrations and used for developing an estimation model can generate small nominal values of mean absolute error and root mean squared error and be translated in high relative levels, which would drive doubts about the suggested estimation model. In this sense, as the predominant difference concerning the works above reported, the present study used as an evaluation metric measure of performance of the proposed estimate model the uncertainty value laid down in the current European Legislation for modeling $PM_{2.5}$ particle levels in ambient air (established in 50% according to the Directive 2008/50/EC). On the other hand, the previous investigation works are limited to a local scope. They do not test the recommended models in other geographical territories, which results in fundamental within the validation process of those predictive models. They do not provide mathematical formulas for achieving this goal. As another distinct, the model proposed in the present work is applied to another geographic area to elucidate whether the implementation of the elaborated model is adequate, not only at the local level but also in a different geographical area.

2.2. *Testing the proposed model to estimate ambient air $PM_{2.5}$ concentrations*

To compare the meteorological scenery that occurred in 2021 concerning that one in 2017-2020, seasonal average 2021 values for each predictor variable are reported. In the case of MSLP, an average value of 1015.72 hPa (1008.52 and 1019.14 hPa minimum and maximum, respectively) was calculated. For the rest of the variables (PBLH, T, P, WD and WS in this order), values of 650.39 m (433.14-767.25 m), 10.28°C (3.15-17.31°C), 2.62 mm/day (1.43-5.74 mm/day), 140.46 degrees (89.20-206.64 degrees) and 4.03 m/sec (2.19-7.14 m/sec) were averaged. At compared the average 2017-2020 vs 2021 values, a percentage variation lower than 10% was observed for each predictor variable. Breaking down the above information for each target study surface (Appendix A Fig. S3), average 2021 values were reasonably similar to those surveyed in the 2017-2020 period, with relative oscillations lower than 10%, except for T (-25.06% in Midlands) and WS (-15.60% in Northlands).

For 2021, the actual average $PM_{2.5}$ concentration, as a function of the target study surface was 6.90 ± 1.65 μg/m^3, 8.42 ± 2.18 μg/m^3 and 9.00 ± 1.99 μg/m^3 for the North-, Mid- and Southlands, respectively. With the proposed mathematical equations, the estimated average $PM_{2.5}$ level was 7.34 ± 2.11 μg/m^3 (Northlands), 8.22 ± 2.87 μg/m^3 (Midlands) and 9.20 ± 1.78 μg/m^3 (Southlands). The execution of an ANOVA test evidenced the difference between current and estimated average $PM_{2.5}$ concentrations was not significant ($p = 0.089$ for Northlands, 0.654 for Midlands and 0.598 for Southlands).

The calculated statistical indicators were: 1.62 μg/m^3, 2.82 μg/m^3 and 20.63% for *RMSE*, *MAE* and *MAPE*, respectively (in Northlands), 1.43 μg/m^3, 2.02 μg/m^3 and 13.80% (Midlands)

and 1.80 μg/m^3, 3.24 μg/m^3 and 16.34% (Southlands). When considering the three covered territorial areas, the highest level of dispersion corresponding to residual values is 1.80 μg $PM_{2.5}$/m^3, sustained by max RMSE value, the highest mean differences between predicted and current values is 3.24 μg $PM_{2.5}$/m^3 (by MAE), and maximum relative error of 20.63% expressed as MAPE, allowing comparing outcomes against other models. In this sense, a relevant investigation study proposed a LUR-GBM model based on land-use regression, the kriging method and light gradient boosting machine to estimate ambient air $PM_{2.5}$ concentrations in China during 2016-2021 (Dai et al., 2022). They obtained significant higher *RMSE* and *MAE* values that those found in the present research.

Fig. 3 shows the outcomes obtained by applying the linear regression test as a function of the three target study surfaces. The current $PM_{2.5}$ concentration explains 57%, 69% and 41% of the total variance of estimated $PM_{2.5}$ concentration for the North-, Mid- and South-lands, respectively. The Durbin–Watson statistic was equal to 1.586 (Northlands), 1.694 (Midlands) and 1.605 (Southlands) and thereby assuming the independence among residues. Considering the evaluation requirements listed in Section 2.4, the simple linear regression equation's ANOVA statistical analysis displayed that the predictive ambient air $PM_{2.5}$ model significantly improves the forecasting of the dependent variable with an F-value of 166.30 (Northlands), 83.38 (Midlands) and 94.16 (Southlands) ($p = 0.00$ in the three cases). Alike, ANOVA of the current $PM_{2.5}$ concentration's coefficient of regression exhibited their significance and incorporation in the linear regression equation ($p = 0.00$ for each target surface).

Based on the last requirement, the implementation of paired samples t-test between estimated and current $PM_{2.5}$ levels exhibited non-significant differences ($p = 0.198$, 0.111 and 0.311 for the Northlands, Midlands and Southlands, respectively).

It also is remarkable to highlight that the parameters defining the regression equation illustrated a slope (0.96, 0.98 and 0.57 for each target surface) and intercept value (1.19, 0.08 and 2.66) close to 1 and 0, respectively. Based on the previously reported findings, it can be concluded that the linear dependence between both target variables is statistically significant, underwriting the applicability of the suggested predictive approach to estimate ambient air $PM_{2.5}$ levels in the function of the surface meteorological conditions. In this sense, significant differences were not found at the seasonal level, given that the current and estimated average concentration of $PM_{2.5}$ particles was 6.90 ± 1.59 vs 7.64 ± 2.02 μg/m^3 (Northlands), 8.47 ± 2.18 vs 8.42 ± 2.59 μg/m^3 (Midlands) and 9.00 ± 1.99 vs 9.80 ± 1.78 μg/m^3 (Southlands).

The current European Legislation lays down a data quality objective for modeling the annual average $PM_{2.5}$ particle concentration in ambient air of 50% expressed as an uncertainty value (Directive 2008/50/EC, annex I). Although the approach recommended in this study covers seasonal periods, this legislative value may offer a view concerning the validity of reached mathematical models. Fig. 3 shows that the modeling dispersion is lower than 50% legislative for each target surface, which provides reliability to the suggested models. Within the same frame, an annual study is addressed. To achieve this objective, all seasonal periods were treated as a

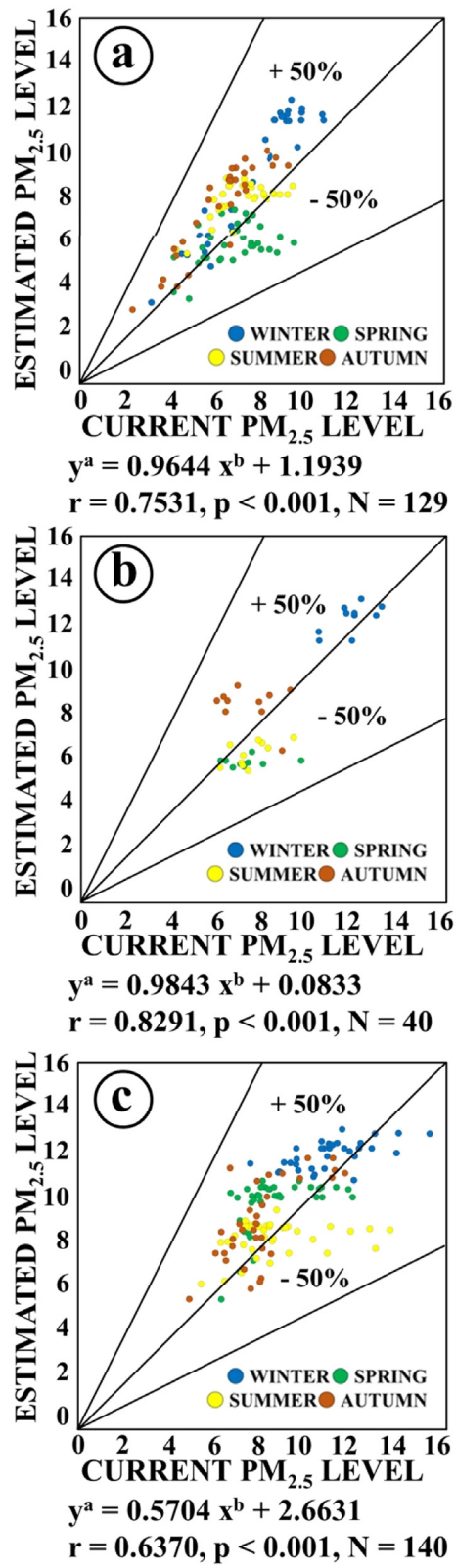

Fig. 3 – Estimated vs current $PM_{2.5}$ concentration (in μg/m^3) along the 2021 year. Key: [a] and [b] correspond to estimated and current concentration, respectively. Note: Graphic a, b and c equal to the Northlands, Midlands and Southlands, respectively.

whole set. In this sense, the current $PM_{2.5}$ concentration accounts for 57% of the total variance of estimated $PM_{2.5}$ concentration (regression equation: Estimated $PM_{2.5} = 0.79$, Current $PM_{2.5} + 1.43$, $r = 0.753$, $p < 0.001$ and $N = 309$). Note that the slope and intercept values were close to 1 and 0, respectively. Non-significant differences were found at the annual level, based on the current vs estimated average concentration of $PM_{2.5}$ particles: 8.05 ± 2.13 μg/m^3 and 8.79 ± 2.27 μg/m^3, respectively. The annual mathematical formulae to estimate $PM_{2.5}$ levels from predictor variables is shown by the following equation:

$$PM_{2.5E} = -240.715 + 0.252 \times MSLP - 0.006 \times PBLH - 0.767 \times P - 0.327 \times T + 0.027 \times WD + 0.258 \times WS \quad (5)$$

The outcomes of the assessment parameters of the model, according to Section 2.4, were as follows: (i) Durbin–Watson statistic: 1.597, guaranteeing the independence among residues; (ii) ANOVA statistical analysis of the simple linear regression equation evidenced that the predictive model significantly enhances the forecasting of the estimated $PM_{2.5}$ concentration with a F-value of 123.78 ($p = 0.000$); and (iii) t-test between estimated and current $PM_{2.5}$ levels showed non-significant differences ($p = 0.068$).

Aspects such as local emission sources, meteorological conditions and topographical features are specific to each geographical zone. This perspective might represent a challenge for the suggested approach in this study. However, the predictor variables are also influenced by these factors, changing ambient air $PM_{2.5}$ levels accordingly. To test this hypothesis, the proposed approach was applied to another air quality network located in another geographical zone with different emission air pollutant focuses, meteorological conditions and topographical parameters. In concrete, two fixed monitoring stations belonging to Madrid City's air quality network were considered. These stations are catalogued as suburban stations (Casa de Campo, Code: 28079024, Latitude: 40°25′09.68′′N, Longitude: 3°44′50.45′′W) and urban traffic stations (Plaza Elíptica, Code: 28079056, Latitude: 40°23′05.87′′N, Longitude: 3°43′07.42′′W). Therefore, these stations present distinct polluting profiles. Current $PM_{2.5}$ concentrations and surface meteorological values for each station were downloaded from Madrid City's open data portal (http://www.mambiente.munimadrid.es). Seasonal average meteorological variables values measured by the stations sited in Casa de Campo and Plaza Elíptica were used to apply the global meteorological formulae previously indicated. The outcomes showed that the $PM_{2.5}$ concentrations in ambient air are efficiently predicted (Fig. 4).

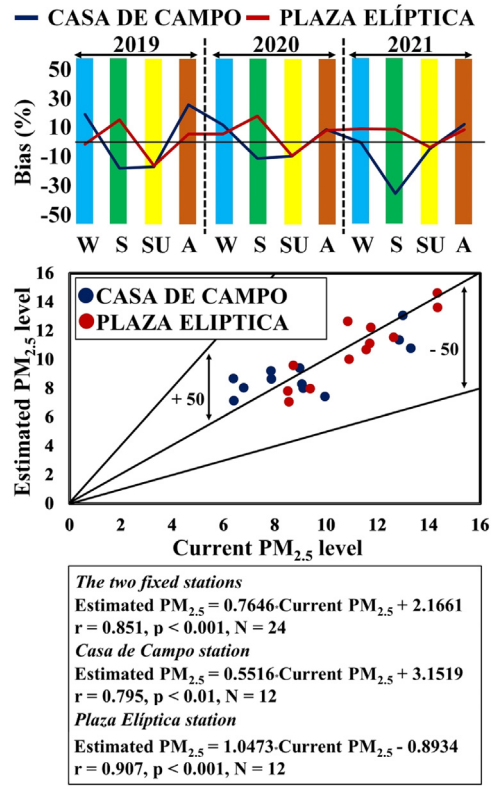

Fig. 4 – Seasonal average estimated and current $PM_{2.5}$ concentration. Top graphic: It represents the bias associated with the estimated $PM_{2.5}$ levels concerning current levels. Lower graphic: It draws the simple linear regression analysis between both variables. Note: W equals to Winter, S to Spring, SU to Summer and A to Autumn.

2.3. *Retrospective analysis of the predictor variables' time series data streams*

A preterit study of surfaced meteorological features' time series was conducted to get valuable past values of these parameters from the last year and seasonal period with data availability up to 2000.

The seasonal average estimated and current value for each predictor variable was correlated once finished the retrospective analysis to evaluate the reliability of the study (Table 2). There is a notable relationship between the seasonal average estimated vs current value for each predictor variable (predictive capacity higher than 52% is reached by executing the retrospective analysis of the target surface meteorological conditions). The average biases (± standard deviation) are reported to be -1.46 × 10^{-5} (±0.32%), -0.95 (±9.58%), -6.82 (±19.38%), -1.63 (±14.88%), 2.57 (±17.41%) and 0.91 (±11.23%) for MSLP, PBLH, P, T, WD and WS, respectively, which further visualize the connection level between both parameters (actual vs estimated). Note that the bias associated for each predictor variable is dragged in the $PM_{2.5}$ estimate process.

2.4. *Spatial variation of $PM_{2.5}$ concentrations in the United Kingdom*

Fig. 5 indicates the spatial distribution of annual average $PM_{2.5}$ concentration in the UK from 2000 to 2021. Between stations, the highest recored concentrations were in midlands and southlands while the lowest levels were in northlands. These elevated observed $PM_{2.5}$ concentrations over the aforementioned areas are likely caused by the emissions (mainly secondary PM precursors) within continental Europe (Broomandi et al., 2021; Vieno et al., 2016). Some of secondary PM precursors are inorganic and organic compo-

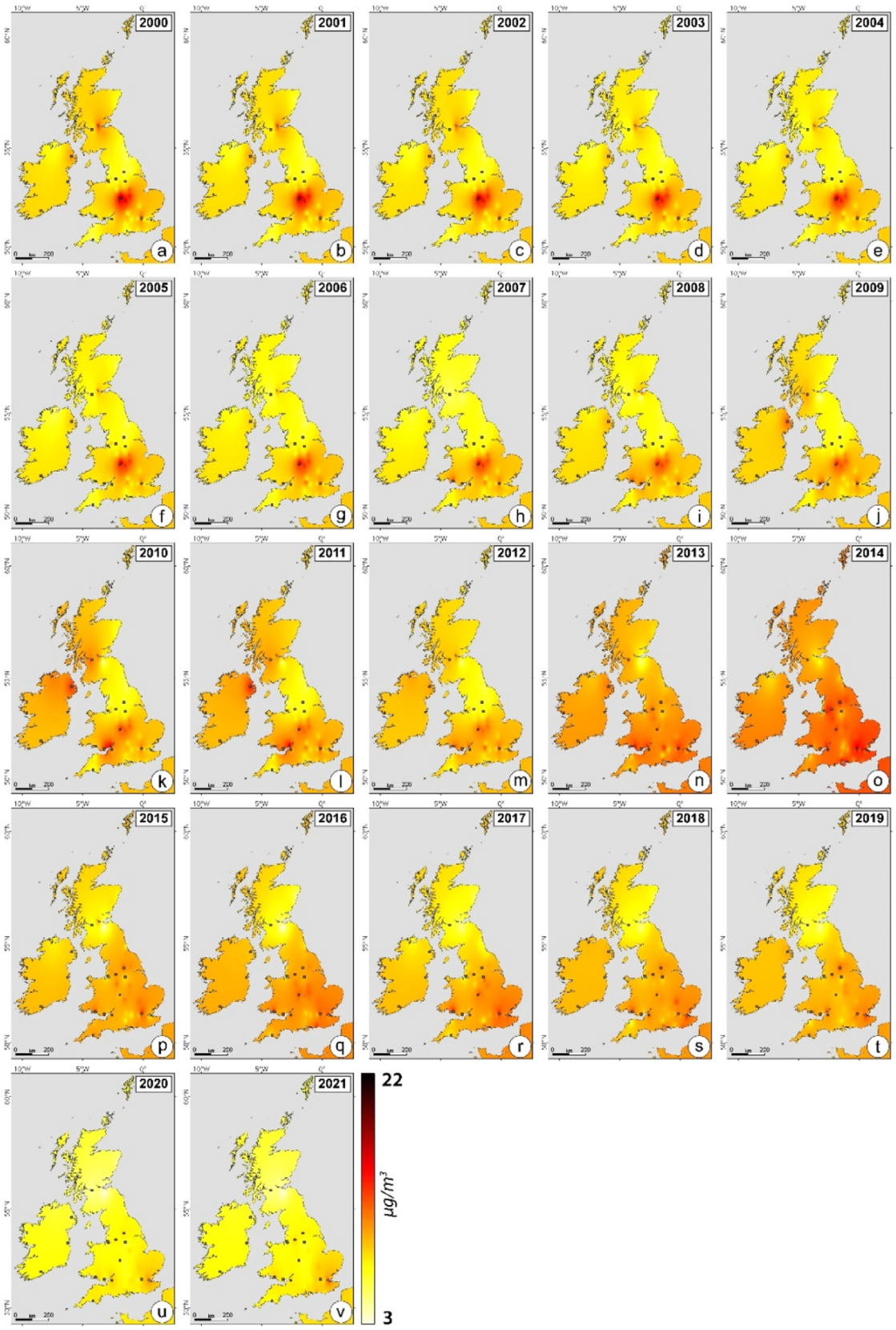

Fig. 5 – Spatial distribution of annual average $PM_{2.5}$ concentration (μg/m³) in the UK from 2000 to 2021.

Table 2 – Outcomes reached by applying a retrospective analysis of the predictor variables' time series data streams.

Predictor variable	Average (SD)		
	Current	Estimated	r^a (CI[b] 95%)
MSLP (kPa)	1,014.3 (3.8)	1,014.4 (2.0)	0.527 (0.501-0.553)
PBLH (m)	686.9 (86.1)	687.5 (59.9)	0.675 (0.655-0.695)
P (mm/day)	2.4 (0.8)	2.4 (0.5)	0.669 (0.649-0.689)
T (°C)	10.4 (3.9)	10.4 (3.8)	0.968 (0.966-0.970)
WD (degrees)	148.4 (31.4)	148.2 (20.9)	0.661 (0.640-0.681)
WS (m/sec)	4.5 (1.2)	4.5 (1.1)	0.889 (0.881-0.896)

Key: [a] Pearson's coefficient of correlation [b] confidence interval.

nents formed within the atmosphere from gaseous precursor emissions, specifically NH_3, NO_x (NO and NO_2), SO_2, and a wide range of non-methane volatile organic compounds (VOC) (Broomandi et al., 2021; Vieno et al., 2016).

The highest $PM_{2.5}$ concentrations observed in northern stations are mainly associated with the winds blowing from the northeast through the east while, those in southern parts are associated with the winds blowing from east through to southeast. Easterly winds can transport emissions (mainly secondary PM precursors) from eastern Europe, northern Germany, Belgium, and Netherlands to the southern parts in the UK. But winds blowing northeast will not pass over the same emission sources, hence the observed concentrations in northern stations with are associated with the north-easterly winds will be lower. Therefore, it can be concluded that the urban background $PM_{2.5}$ concentrations are mainly influenced by regional emission sources rather then local emission sources and PM emissions likely secondary PM precursors originating from continental Europe have an important role in affecting $PM_{2.5}$ levels across the UK (Barry and Chorley, 2010; Harrison et al., 2012). Broomandi et al. (2021) by constructing a directed Granger causality and an undirected Correlation networks, showed two robust spatial interdependent communities divide the United Kingdom into the northern and southern city clusters. They concluded that one compelling reason of connection within the network is having common transboundary sources.

The highest concentrations were observed in springtime and wintertime while the lowest were in summer and fall (Appendix A Fig. S4). The worst dispersion conditions for the $PM_{2.5}$ air pollution level occur at the low amount of precipitation (negative correlation), low PBLH (negative correlation), high pressure atmospheric SLP pattern (positive correlation), low wind speed (negative correlation), and low air temperatures (negative correlation) along with a wind direction from areas with a higher accumulation of pollution sources (Zhou et al., 2007) (Fig. 6). The ambient air temperature accompanied with atmospheric stability affects the vertical dispersion of pollutants but the wind profile influences the horizontal dispersion of pollutants. Unstable stratification causes favourable conditions for pollutant dispersion, but stable weather condition causes a weak vertical mixing condition (Buchanan et al., 2002).

The spring peak is due to the combination of easterlies importing pollution from mainland Europe and mild temperatures enhancing partitioning of gas phase precursors into the aerosol phase, a peak in NH_3 emissions due to March-April fertilizer application, and stagnant conditions over the UK limiting ventilation (Hellsten et al., 2008, Harrison et al., 2012, Vieno et al., 2014, Fenech et al., 2019). In wintertime, the $PM_{2.5}$ peaks could be attributed to the mild unfavourable or unfavourable dispersion conditions along with the higher amount of emissions of both primary and secondary PM.

2.5. *Long-term $PM_{2.5}$ trends in the United Kingdom*

To offer a practical application of the suggested models, a long-time $PM_{2.5}$ particle trend study is performed on the target surface. For that, meteorological variables' data from the retrospective analysis were used to estimate $PM_{2.5}$ concentrations in years not measured by the target fixed monitoring stations. The attended period was from 2000 to 2021. The final result of this study reported current (measured by fixed stations) and estimated $PM_{2.5}$ levels for each monitoring station.

We applied Mann-Kendal test and Sen's slope test on the daily $PM_{2.5}$ data in the UK, which showed that there is a general statistically significant reduction in most of the studied cities in North- Mid-, and Southlands between 2000 and 2021 in springtime (April-June) (Fig. 7-8).

While in wintertime (January-March), studied stations across the UK showed a general statistically significant increasing trend with a small magnitude of changes via Sen's slope estimator. Moreover, $PM_{2.5}$ concentration followed a nearly increasing trend in summer (July-September) and autumn (October-December) (p-values below 0.05) in the UK but not in the Mainland stations.

The observed increases in the $PM_{2.5}$ concentration could be attributed to the possible increase in the transboundary emissions likely secondary PM precursors originating from continental Europe and agricultural emissions of NH_3 in the UK. Previous studies indicated that potentially important $PM_{2.5}$ sources are national agricultural emissions of NH_3 (Vieno et al., 2016), global shipping fleet (Jonson et al., 2020), and anthropogenic emissions from neighbouring countries in continental Europe (Vieno et al., 2014), transported dust from North Africa (Wang et al., 2020) and from uncertain anthropogenic sources (Philip et al., 2017) in the UK.

Since 2010, NO_x and SO_2 emissions have declined by 33% and 60%, but NH_3 emissions have increased by 25% (Richmond et al., 2020). Only agriculture dominates (~ 90%) anthropogenic NH_3 emissions in the UK (Marais et al., 2021). Kelly et al. (2023) showed that city sources had a smaller por-

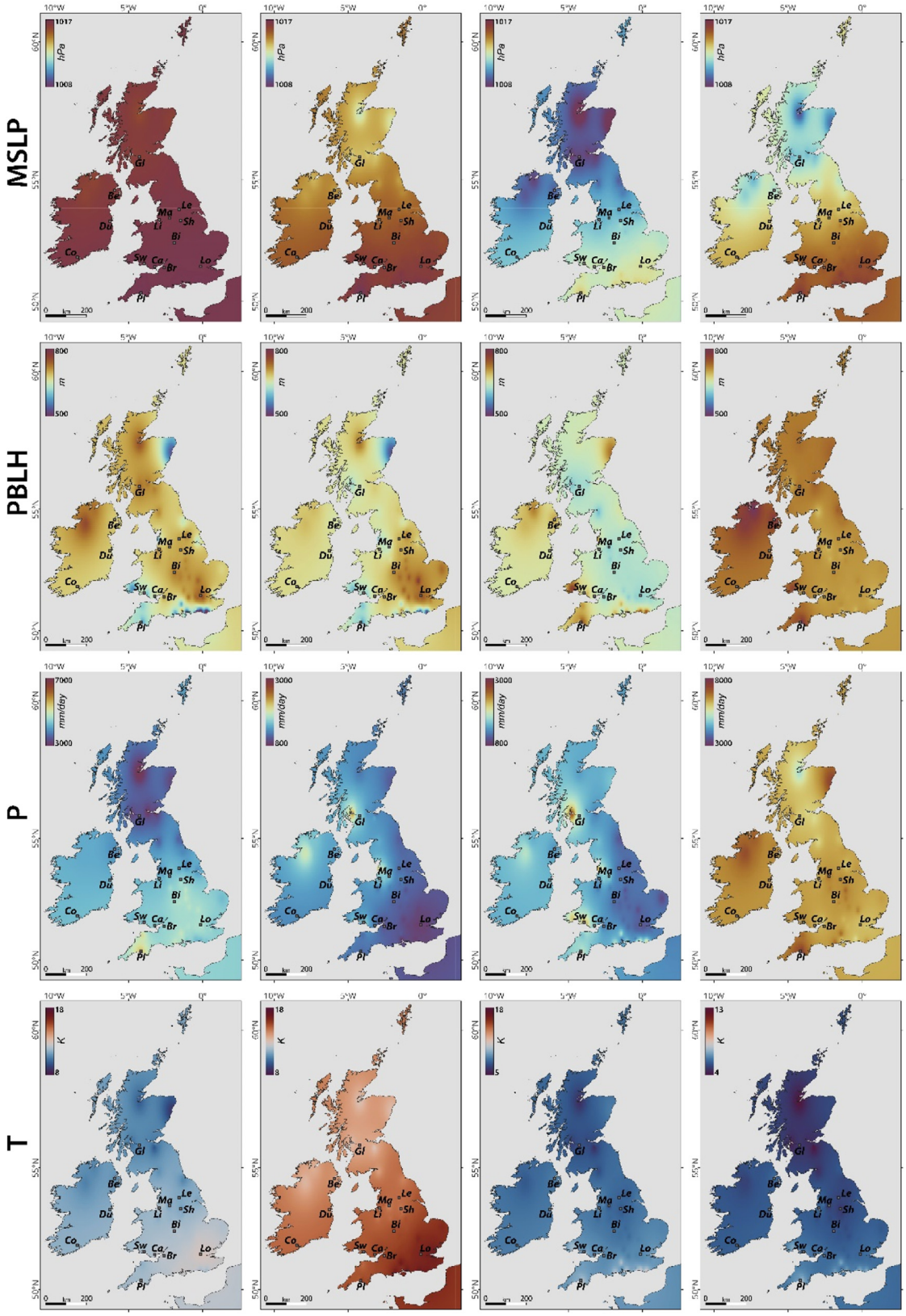

Fig. 6 – Governing meteorological parameters over UK between 2000 and 2021 in different seasons. Note: MSLP corresponds to the mean sea-level pressure, PBLH to the planetary boundary layer height, T to the temperature, and P to the precipitation.

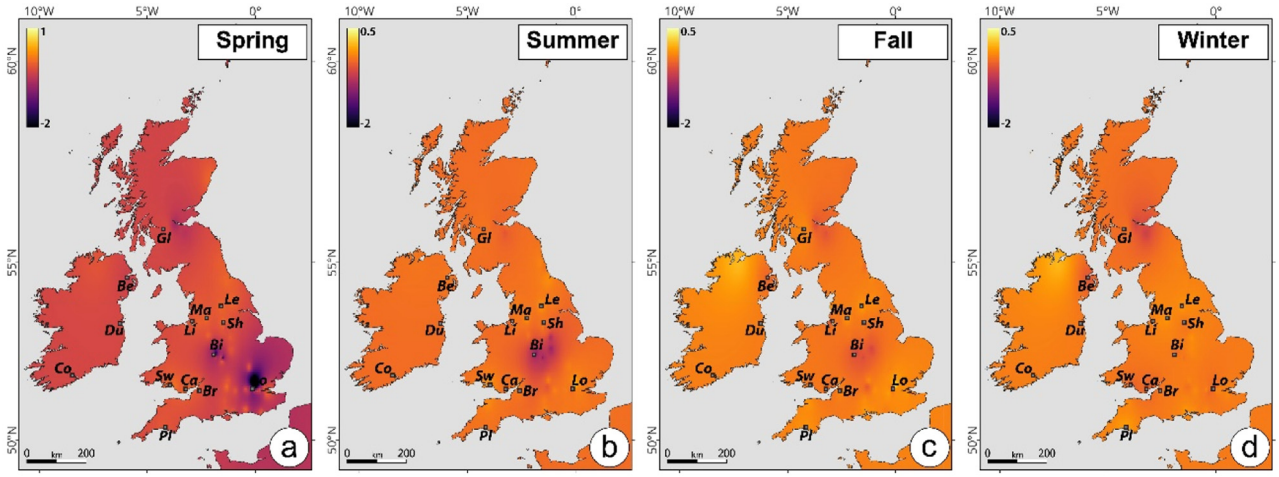

Fig. 7 – Observed trend of changes in $PM_{2.5}$ concentration using Sen's Slope estimator in (a) Spring, (b) Summer, (c) Fall, and (d) Winter across the UK from 2000 to 2021.

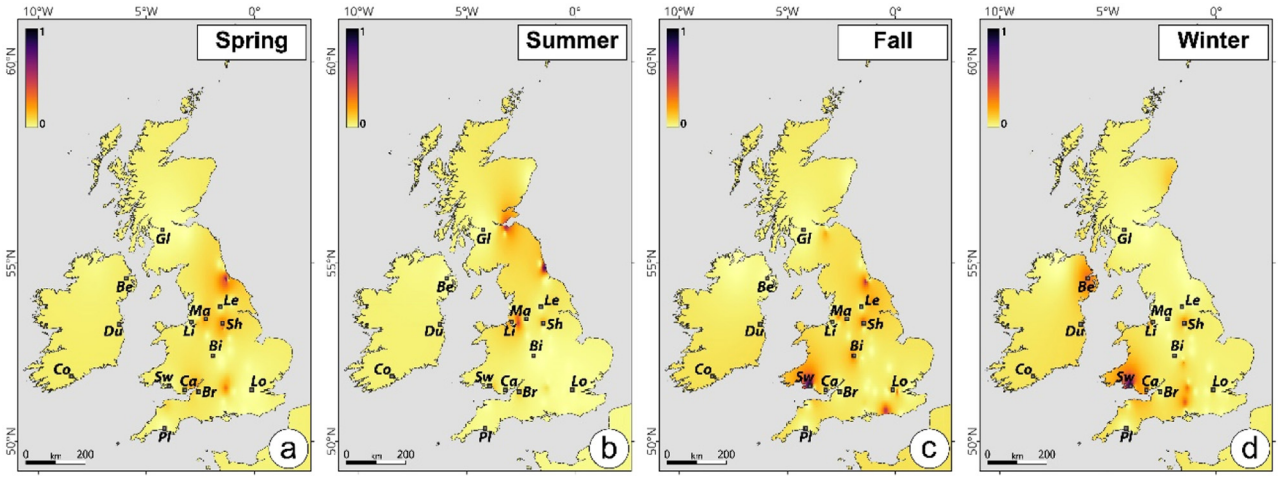

Fig. 8 – Significance (*p*-value) of changes in $PM_{2.5}$ concentration using Sen's Slope estimator in (a) Spring, (b) Summer, (c) Fall, and (d) Winter across the UK from 2000 to 2021.

tion of annual mean $PM_{2.5}$ compared to agricultural emissions from NH_3 and continental Europe. The dominant contribution of agricultural emissions of NH_3 to $PM_{2.5}$ pollution is also reported in similar studies at regional and global scales (Lelieveld et al., 2015; Pozzer et al., 2017; Gu et al., 2021). Therefore, actions at the national level are required to control and/or reduce the agricultural sources on NH_3, which accounts for 25%-39% of urban $PM_{2.5}$ pollution. Moreover, supranational agreements should be strengthened to effectively alleviate the $PM_{2.5}$ level in vast parts of the UK by reducing long-range transport of pollution from continental Europe which accounts for 16%-28% of urban $PM_{2.5}$ pollution.

3. Conclusion

The findings presented in this research work support the feasibility of an easily applicable approach for estimating ambient air $PM_{2.5}$ concentrations using various meteorological parameters, such as mean sea-level pressure, planetary boundary layer height, temperature, precipitation, wind direction, and speed. These meteorological features are more widely available compared to air pollutant data. By utilizing this approach, the model can estimate $PM_{2.5}$ levels in geographic areas where monitoring is not conducted, thereby providing valuable supplementary data on $PM_{2.5}$ exposure levels and spatial-temporal patterns of pollution.

The United Kingdom was chosen as a case study to develop and test the proposed model. Despite differences in latitudes, longitudes, emission sources, and dispersion characteristics between the original study area and the secondary area, the results support the implementation of the mathematical model to estimate $PM_{2.5}$ concentrations in any geographic zone using the relevant meteorological data. This approach is particularly relevant as meteorological monitoring networks tend to have a larger number of stations compared to air quality networks, thereby offering additional $PM_{2.5}$ data without the need for establishing new monitoring sites.

Potential applications of this model include assisting competent authorities in decision-making regarding $PM_{2.5}$ control measures, providing scientific support for European legislative guidelines, enhancing understanding of ambient air $PM_{2.5}$ behavior, offering comprehensive information on spatial distribution gradients of $PM_{2.5}$, and facilitating epidemiological-based studies for risk assessment.

However, it should be noted that the proposed model should be applied in geographic areas where the necessary meteorological variables are measured. As a future step, the recommended methodological framework could be extended to include PM_{10}, as it also has significant impacts on human health.

Data availability

Data is available on request from the authors.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

This work was supported by the Collaborative Research Project (CRP) grant, Nazarbayev University (Nos. 11022021CRP1512, 211123CRP1604). PK acknowledges the support from the NERC-funded projects ASAP-Delhi (NE/P016510/1), GreenCities (NE/P016510/1) and RECLAIM Network Plus (EP/W034034/1).

Supplementary materials

Supplementary material associated with this article can be found in the online version at doi:10.1016/j.jes.2023.11.019.

References

Annesi-Maesano, I., Forastiere, F., Kunzli, N., Brunekref, B., 2007. Particulate matter, science and EU policy. Eur. Respir. J. 29 (3), 428–431. doi:10.1183/09031936.00129506.

Barman, U., Choudhury, R.D., 2020. Smartphone image based digital chlorophyll meter to estimate the value of citrus leaves chlorophyll using linear regression, LMBP-ANN and SCGBP-ANN. J. King Saud Univ. Comp. Info. Sci. 34 (6), 2938–2950. doi:10.1016/j.jksuci.2020.01.005.

Barry, R.G., Chorley, R.J., 2010. Atmosphere, Weather and Climate, ninth ed. ed. Routledge, Abingdon.

Bi, J., Belle, J.H., Wang, Y., Lyapustin, A.I., Wildani, A., Liu, Y., 2019. Impacts of snow and cloud covers on satellite-derived PM2.5 levels. Remote Sens. Environ. 221, 665–674. doi:10.1016/j.rse.2018.12.002.

Broomandi, P., Geng, X., Guo, W., Pagani, A., Topping, D., Kim, J.R., 2021. Dynamic complex network analysis of PM2.5 concentrations in the UK, using hierarchical directed graphs (V1.0.0). Sustainability 13 (4), 2201. doi:10.3390/su13042201.

Buchanan, C.M., Beverland, I.J., Heal, M.R., 2002. The influence of weather-type and long-range transport on airborne particle concentrations in Edinburgh, UK. Atmos. Environ. 36 (34), 5343–5354. doi:10.1016/S1352-2310(02)00579-4.

Chen, Z.Y., Zhang, T.H., Zhang, R., Zhu, Z.M., Yang, J., Chen, P.Y., et al., 2019. Extreme gradient boosting model to estimate PM2.5 concentrations with missing-filled satellite data in China. Atmos. Environ. 202, 180–189. doi:10.1016/j.atmosenv.2019.01.027.

Chu, Y., Liu, Y., Liu, Z., Lu, H., Lu, Y., Mao, Z., et al., 2016. A review on predicting ground PM2.5 concentration using satellite aerosol optical depth. Atmosphere (Basel) 7 (10). doi:10.3390/atmos7100129.

COMEAP, 2010. The mortality effects of long-term exposure to particulate air pollution in the United Kingdom.

Cortina-Januchs, M.G., Quintanilla-Dominguez, J., Vega-Corona, A., Andina, D., 2015. Development of a model for forecasting of PM10 concentrations in salamanca, Mexico. Atmos. Pollut. Res. 6, 626–634. doi:10.5094/APR.2015.071.

Crumeyrolle, S., Chen, G., Ziemba, L., Beyersdorf, A., Thornhill, L., Winstead, E., et al., 2014. Factors that influence surface PM2.5 values inferred from satellite observations: perspective gained for the Baltimore-Washington area during DISCOVER-AQ. Atmos. Chem. Phys. 14, 2139–2153. doi:10.5194/acp-14-2139-2014.

Dai, H., Huang, G., Wang, J., Zeng, H., Zhou, F., 2022. Spatio-Temporal characteristics of PM2.5 concentrations in China based on multiple sources of data and LUR-GBM during 2016-2021. Int. J. Environ. Res. Public Health 19, 6292. doi:10.3390/ijerph19106292.

DEFRA, 2015. Department for Environment, Food and Rural Affairs. United Kingdom.

Doreswamy, Harishkumar, K.S., Yogesh, K.M., Gad, I., 2020. Forecasting air pollution particulate matter (PM2.5) using machine learning regression models. Procedia Comput. Sci. 171, 2057–2066. doi:10.1016/j.procs.2020.04.221.

Duncan, B., Prados, A., Lamsal, L., Liu, Y., Streets, D., Gupta, P., et al., 2014. Satellite data of atmospheric pollution for U.S. air quality applications: examples of applications, summary of data end-user resources, answers to FAQs, and common mistakes to avoid. Atmos. Environ. 94, 647–662. doi:10.1016/j.atmosenv.2014.05.061.

EEA, 2014. Air quality in Europe –2014 report. EEA Report No 5/2014.

ERA5, 2020. ERA 5 global climate reanalysis [WWW Document].

Eurachem, 2012. Quantifying uncertainty in analytical measurement. EURACHEM /CITAC Guide CG 4 3rd ed..

Fenech, S., Doherty, R.M., Heaviside, C., Macintyre, H.L., O'Connor F., M., Vardoulakis, S., et al., 2019. Meteorological drivers and mortality associated with O3 and PM2.5 air pollution episodes in the UK in 2006. Atmos. Environ. 213, 699–710. https://doi:10.1016/j.atmosenv.2019.06.030.

Galán-Madruga, D., 2022. Urban air quality changes resulting from the lockdown period due to the COVID19 pandemic. Int. J. Environ. Sci. Technol. 20, 7083–7098. doi:10.1007/s13762-022-04464-6, Published online August.

Galán-Madruga, D., García Cambero, J.P., 2022. An optimized approach for estimating benzene in ambient air within an air quality monitoring network. J. Environ. Sci. 111, 164–174. doi:10.1016/j.jes.2021.03.005.

Gehrig, R., Buchmann, B., 2003. Characterising seasonal variations and spatial distribution of ambient PM10 and PM2.5 concentrations based on long-term Swiss monitoring data. Atmos. Environ. 37, 2571–2580. doi:10.1016/S1352-2310(03)00221-8.

Gehring, U., Wijga, A.H., Hoek, G., Bellander, T., Berdel, D., Brüske, et al., 2015. Exposure to air pollution and development of asthma and rhinoconjunctivitis throughout childhood and

adolescence: a population-based birth cohort study. Lancet. Respir. Med. 3, 933–942. doi:10.1016/S2213-2600(15)00426-9.

Geng, G., Murray, N.L., Chang, H.H., Liu, Y., 2018. The sensitivity of satellite-based PM2.5 estimates to its inputs: implications to model development in data-poor regions. Environ. Int. 121, 550–560. doi:10.1016/j.envint.2018.09.051.

Geng, G., Meng, X., He, K., Liu, Y., 2020. Random forest models for PM2.5 speciation concentrations using MISR fractional AODs. Environ. Res. Lett. 15, 34056. doi:10.1088/1748-9326/ab76df.

Gowers, A.M., Miller, B.G., Stedman, J.R., 2014. Estimating local mortality burdens associated with particulate air pollution © Crown copyright 2014, licenced under the Open Government Licence (OGL).

Gu, B., Zhang, L., Van Dingenen, R., Vieno, M., Van Grinsven, H.J.M., Zhang, X., 2021. Abating ammonia is more cost-effective than nitrogen oxides for mitigating PM2.5 air pollution. Science 374, 758–762. https://doi:10.1126/science.abf8623.

Guevara-Luna, M., Guevara Luna, F., Mendez Espinosa, J., Belalcazar, L., 2018. Spatial and temporal assessment of particulate matter using AOD data from MODIS and surface measurements in the ambient air of Colombia. Asian J. Atmos. Environ. 12, 165. doi:10.5572/ajae.2018.12.2.165.

Harrison, R.M., Laxen, D., Moorcroft, S., Laxen, K., 2012. Processes affecting concentrations of fine particulate matter (PM2.5) in the UK atmosphere. Atmos. Environ. 46, 115–124. doi:10.1016/J.ATMOSENV.2011.10.028.

Heal, MR, Kumar, P, Harrison, RM., 2012. Particles, air quality, policy and health. Chem Soc Rev. 7 41 (19), 6606–6630. https://doi:10.1039/c2cs35076a. Epub 2012 Jun 1. PMID: 22660420.

Hellsten, S., Dragosits, U., Place, C.J., Vieno, M., Dore, A.J., Misselbrooket, T.H., et al., 2008. Modelling the spatial distribution of ammonia emissions in the UK. Environ. Pollut. 154, 370–379. https://doi:10.1016/j.envpol.2008.02.017.

Huang, T., Yu, Y., Wang, H., Huang, W., Chen, X., 2018. Spatial–seasonal characteristics and critical impact factors of PM2.5 concentration in the Beijing–Tianjin–Hebei urban agglomeration. PLoS One 13, e0201364. doi:10.1371/journal.pone.0201364.

Jiang, Q., Christakos, G., 2018. Space-time mapping of ground-level PM2.5 and NO2 concentrations in heavily polluted northern China during winter using the Bayesian maximum entropy technique with satellite data. Air Qual. Atmos. Heal. 11, 23–33. doi:10.1007/s11869-017-0514-8.

Jonson, J.E., Gauss, M., Schulz, M., Jalkanen, J.P., Fagerli, H., 2020. Effects of global ship emissions on European air pollution levels. Atmospheric Chem. Phys. 20, 11399–11422. https://doi:10.5194/acp-20-11399-2020.

Karunasingha, D.S.K., 2022. Root mean square error or mean absolute error? Use their ratio as well. Inf. Sci. 585, 609–629. doi:10.1016/j.ins.2021.11.036.

Kelly, L.M., Marais, E.A., Lu, G., Obszynska, J., Mace, M., White, J., et al., 2023. Diagnosing domestic and transboundary sources of fine particulate matter (PM2.5) in UK cities using GEOSChem. City Environ. Interact. 18, 100100. doi:10.1016/j.cacint.2023.100100.

Kottek, M., Grieser, J., Beck, C., Rudolf, B., Rubel, F., 2006. World map of the Köppen-Geiger climate classification updated. Meteorol. Zeitschrift 15, 259–263. doi:10.1127/0941-2948/2006/0130.

Krishna, R.K., Ghude, S.D., Kumar, R., Beig, G., Kulkarni, R., Nivdange, S., et al., 2019. Surface PM2.5 Estimate Using Satellite-Derived Aerosol Optical Depth over India. Aerosol Air Qual. Res. 19, 25–37. doi:10.4209/aaqr.2017.12.0568.

Kumar, A., Goyal, P., 2011. Forecasting of air quality in Delhi using principal component regression technique. Atmos. Pollut. Res. 2, 436–444. doi:10.5094/APR.2011.050.

Kumar, S., Mishra, S., Singh, S.K., 2020. A machine learning-based model to estimate PM2.5 concentration levels in Delhi's atmosphere. Heliyon 6 (11), e05618. doi:10.1016/j.heliyon.2020.e05618.

Lelieveld, J., Evans, J.S., Fnais, M., Giannadaki, D., Pozzer, A., 2015. The contribution of outdoor air pollution sources to premature mortality on a global scale. Nature 525, 367–371. https://doi:10.1038/nature15371.

Lim, S., Vos, T., Flaxman, A., Danaei, G., Shibuya, K., Adair-Rohani, H., et al., 2012. A comparative risk assessment of burden of disease and injury attributable to 67 risk factors and risk factor clusters in 21 regions, 1990-2010: a systematic analysis for the Global Burden of Disease Study 2010. Lancet 380, 2224–2260. doi:10.1016/S0140-6736(12)61766-8.

Lin, C., Li, Y., Yuan, Z., Lau, A., Li, C., Fung, J., 2015. Using satellite remote sensing data to estimate the high-resolution distribution of ground-level PM2.5. Remote Sens. Environ 156, 117–128. doi:10.1016/j.rse.2014.09.015.

Liu, Y., Koutrakis, P., Kahn, R., Turquety, S., Yantosca, R., 2007. Estimating fine particulate matter component concentrations and size distributions using satellite-retrieved fractional aerosol optical depth: part 2—a case study. J. Air Waste Manag. Assoc. 57, 1360–1369. doi:10.3155/1047-3289.57.11.1360.

Ma, Z., Hu, X., Huang, L., Bi, J., Liu, Y., 2014. Estimating ground-level PM2.5 in China using satellite remote sensing. Environ. Sci. Technol. 48, 7436–7444. doi:10.1021/es5009399.

Marais, E.A., Pandey, A.K., Van Damme, M., Clarisse, L., Coheur, P.F., Shephard, M.W., et al., 2021. UK ammonia emissions estimated with satellite observations and GEOS-Chem. J. Geophys. Res. Atmos. 126 (18), e2021JD035237. https://doi:10.1029/2021jd035237.

Masood, A., Ahmad, K., 2020. A model for particulate matter (PM2.5) prediction for Delhi based on machine learning approaches. Procedia Comput. Sci. 167, 2101–2110. doi:10.1016/j.procs.2020.03.258.

Nabavi, S.O., Haimberger, L., Abbasi, E., 2018. Assessing PM2.5 concentrations in Tehran, Iran, from space using MAIAC, deep blue, and dark target AOD and machine learning algorithms. Atmos. Pollut. Res. 10 (3), 889–903. doi:10.1016/j.apr.2018.12.017.

ONS, 2020. Office for National Statistics (ONS).

Peng, J., Han, H., Yi, Y., Huang, H., Xie, L., 2022. Machine learning and deep learning modeling and simulation for predicting PM2.5 concentrations. Chemosphere 308, 136353. doi:10.1016/j.chemosphere.2022.136353.

Philip, S., Martin, R.V., Snider, G., Weagle, C.L., Donkelaar, A.V., Brauer, M., et al., 2017. Anthropogenic fugitive, combustion and industrial dust is a significant, underrepresented fine particulate matter source in global atmospheric models. Environ. Res. Lett. 12 (4), 044018. https://doi:10.1088/1748-9326/aa65a4.

Pozzer, A., Tsimpidi, A.P., Karydis, V.A., de Meij, A., Lelieveld, J., 2017. Impact of agricultural emission reductions on fine-particulate matter and public health. Atmospheric Chem. Phys. 17, 12813–12826. https://doi:10.5194/acp-17-12813-2017.

Ramsey, N.R., Klein, P.M., Moore, B., 2014. The impact of meteorological parameters on urban air quality. Atmos. Environ. 86, 58–67. doi:10.5094/APR.2011.050.

Richmond, B., Misra, A., Brown, P., Karagianni, E., Murrells, T., Pang, Y., 2020. UK informative inventory report (1990 to 2018). Available: https://ukair.defra.gov.uk/assets/documents/reports/cat07/2003131327_GB_IIR_2020_v1.0.pdf [Accessed 11-04- 2022].

Schneider dos Santos, R., Vicedo-Cabrera, A., Sera, F., Stafoggia, M., de Hoogh, K., Kloog, I., et al., 2020. A satellite-based spatio-temporal machine learning model to reconstruct daily PM2.5 concentrations across Great Britain. medRxiv doi:10.1101/2020.07.19.20157396, 2020.07.19.20157396.

Shin, M., Kang, Y., Park, S., Im, J., Yoo, C., Quackenbush, L.J., 2020. Estimating ground-level particulate matter concentrations using satellite-based data: a review. GIScience Remote Sens 57, 174–189. doi:10.1080/15481603.2019.1703288.

Vieno, M., Heal, M., Williams, M., Carnell, E., Nemitz, E., Stedman, R., et al., 2016. The sensitivities of emissions reductions for the mitigation of UK PM2.5. Atmos. Chem. Phys. 16, 265–276. doi:10.5194/acp-16-265-2016.

Vieno, M., Heal, M.R., Hallsworth, S., Famulari, D., Doherty, R.M., Dore, A.J., et al., 2014. The role of long-range transport and domestic emissions in determining atmospheric secondary inorganic particle concentrations across the UK. Atmospheric Chem. Phys. 14, 8435–8447. https://doi:10.5194/acp-14-8435-2014.

Wang, Q.Q., Gu, J.W., Wang, X.R., 2020. The impact of Sahara dust on air quality and public health in European countries. Atmos. Environ. 241, 10. https://doi:10.1016/j.atmosenv.2020.117771.

Wei, J., Huang, W., Li, Z., Xue, W., Peng, Y., Sun, L., et al., 2019. Estimating 1-km-resolution PM2.5 concentrations across China using the space-time random forest approach. Remote Sens. Environ. 231, 111221. doi:10.1016/j.rse.2019.111221.

Wongwatcharapaiboon, J., 2020. Review article: toward future particulate matter situations in Thailand from supporting policy, Network and Economy. Futur. Cities Environ. 6 (1). http://doi.org/10.5334/fce.79.

Xiao, Q., Chang, H.H., Geng, G., Liu, Y., 2018. An ensemble machine-learning model to predict historical PM2.5 concentrations in China from satellite data. Environ. Sci. Technol. 52, 13260–13269. doi:10.1021/acs.est.8b02917.

Xiao, Q., Wang, Y., Chang, H., Meng, X., Geng, G., Lyapustin, A., et al., 2017. Full-coverage high-resolution daily PM 2.5 estimation using MAIAC AOD in the Yangtze River Delta of China. Remote Sens. Environ. 199, 437–446. doi:10.1016/j.rse.2017.07.023.

Xu, X., Zhang, C., 2020. Estimation of ground-level PM2.5 concentration using MODIS AOD and corrected regression model over Beijing, China. PLoS One 15, e0240430.

Yang, Y., Sun, L., Guo, C., 2018. Aero-material consumption prediction based on linear regression model. Procedia Comput. Sci. 131, 825–831. doi:10.1016/j.procs.2018.04.271.

Yazdi, M., Kuang, Z., Dimakopoulou, K., Barratt, B., Suel, E., Amini, H., et al., 2020. Predicting fine particulate matter (PM2.5) in the greater london area: an ensemble approach using machine learning methods. Remote Sens. 12 (6), 914. doi:10.3390/rs12060914.

Zeydan, Ö., Wang, Y., 2019. Using MODIS derived aerosol optical depth to estimate ground-level PM2.5 concentrations over Turkey. Atmos. Pollut. Res. 10, 1565–1576. doi:10.1016/j.apr.2019.05.005.

Zhang, Y., Cao, F., 2015. Fine particulate matter (PM2.5) in China at a city level. Sci. Rep. 5, 14884. doi:10.1038/srep14884.

Zhao, R., Gu, X., Xue, B., Zhang, J., Ren, W., 2018. Short period PM2.5 prediction based on multivariate linear regression model. PLoS One 13, e0201011. doi:10.1371/journal.pone.0201011.

Zhong, J., Zhang, X., Gui, K., Wang, Y., Che, H., Shen, X., et al., 2021. Robust prediction of hourly PM2.5 from meteorological data using LightGBM. Natl. Sci. Rev. 8, nwaa307. doi:10.1093/nsr/nwaa307.

Zhou, K., Ye, Y., Liu, Q., Liu, A., Peng, S., 2007. Evaluation of ambient air quality in Guangzhou, China. *J. Environ. Sci.* (China) 19, 432–437. doi:10.1016/S1001-0742(07)60072-2.