RESEARCH

Developing PM$_{2.5}$ and PM$_{10}$ prediction models on a national and regional scale using open-source remote sensing data

Luka Mamić · **Mateo Gašparović** ·
Gordana Kaplan

Abstract Clean air is the precursor to a healthy life. Air quality is an issue that has been getting under its well-deserved spotlight in the last few years. From a remote sensing point of view, the first Copernicus mission with the main purpose of monitoring the atmosphere and tracking air pollutants, the Sentinel-5P TROPOMI mission, has been widely used worldwide. Particulate matter of a diameter smaller than 2.5 and 10 μm (PM$_{2.5}$ and PM$_{10}$) significantly determines air quality. Still, there are no available satellite sensors that allow us to track them remotely with high accuracy, but only using ground stations. This research aims to estimate PM$_{2.5}$ and PM$_{10}$ using Sentinel-5P and other open-source remote sensing data available on the Google Earth Engine (GEE) platform for heating (December 2021, January, and February 2022) and non-heating seasons (June, July, and August 2021) on the territory of the Republic of Croatia. Ground stations of the National Network for Continuous Air Quality Monitoring were used as a starting point and as ground truth data. Raw hourly data were matched to remote sensing data, and seasonal models were trained at the national and regional scale using machine learning. The proposed approach uses a random forest algorithm with a percentage split of 70% and gives moderate to high accuracy regarding the temporal frame of the data. The mapping gives us visual insight between the ground and remote sensing data and shows the seasonal variations of PM$_{2.5}$ and PM$_{10}$. The results showed that the proposed approach and models could efficiently estimate air quality.

Keywords Air quality · Google Earth Engine · Prediction model · Remote sensing · Sentinel-5P

L. Mamić (✉)
Department of Civil, Building and Environmental
Engineering, Sapienza University of Rome, Rome, Italy
e-mail: luka.mamic@uniroma1.it

L. Mamić
Department of Land, Environment, Agriculture
and Forestry (TESAF), University of Padua, Padova, Italy

M. Gašparović
Chair of Photogrammetry and Remote Sensing, Faculty
of Geodesy, University of Zagreb, Zagreb, Croatia
e-mail: mateo.gasparovic@geof.unizg.hr

G. Kaplan
Institute of Earth and Space Sciences, Eskisehir Technical
University, Eskisehir, Turkey
e-mail: kaplangorde@gmail.com

Introduction

Air pollution is a significant threat to modern society, and it has been shown that it negatively affects both, the people's health (Ghorani-Azam et al., 2016; Lave & Seskin, 2013; Shahriyari et al., 2022) and the environment (Saurabh Sonwani & Vandana Maurya, 2019; Stevens et al., 2020). Particulate matter (PM) with an effective aerodynamic diameter smaller than 2.5 and 10 μm (PM$_{2.5}$ and PM$_{10}$) have gained specific attention among air pollutants, and their effects on human health and ecosystems have

become important research topics in recent years (Bodor et al., 2022; Faraji Ghasemi et al., 2020; Leão et al., 2022; Yunesian et al., 2019; Zoran et al., 2020).

High levels of PM_{10} have been associated with increased hospital admissions for lung and heart disease. In contrast, $PM_{2.5}$ has a greater negative impact on human health than PM_{10} because it penetrates deeper into the respiratory system. Prediction of atmospheric composition aids significantly in air quality management; however, predicting air quality remains a challenge due to the processes' complexity and the strong coupling across many parameters, which affects modeling performance (Biancofiore et al., 2017). Overall, the composition and concentration of all components in PM of all size fractions vary. This variability is most likely caused by natural and environmental factors such as human or natural sources, temperature and seasonal changes, and geographical location (Polichetti et al., 2009).

In Croatia, the National Network for Continuous Air Quality Monitoring uses ground stations to monitor atmospheric concentrations of $PM_{2.5}$ and PM_{10} throughout the country. However, as noted by Li et al. (2020), ground station measurements are only applicable in local areas and cannot provide a broad perspective. Recently, several remote sensing-based methods have been developed for this issue. Estimating ambient $PM_{2.5}$ and PM_{10} concentrations using observations from remote sensing satellites have been the subject of various studies to date (Lin et al., 2015; Zhang & Li, 2015; Chen et al., 2018a, b; Sun et al., 2019; Li et al., 2021). Scientists have also used high-resolution remote sensing imagery from Unmanned Aerial Vehicles (UAV) for computer simulation and comparison with real measurements (Cichowicz & Dobrzański, 2022), 3D investigation of air quality (Samad et al., 2022), pollutants detection (De Fazio et al., 2022), etc. However, the mentioned studies have been conducted over small study areas. For vast areas, the use of Sentinel-5P mission TROPOMI data to estimate $PM_{2.5}$ and PM_{10} concentrations has been implemented by a few studies lately (Ahmed et al., 2022; Han et al., 2022; Li et al., 2022; Son et al., 2022; Wang et al., 2021). Sentinel-5P TROPOMI was launched on October 13, 2017, as the first Copernicus mission with the main objective of monitoring air pollutants in the atmosphere. It is the most recent global satellite mission in monitoring air quality and daily measures concentrations of ozone (O_3), methane (CH_4), formaldehyde (HCHO), carbon monoxide (CO), nitrogen oxide (NO_2), sulphur dioxide (SO_2), and aerosol—provided as an aerosol index (AI).

Wang et al. (2021) used Sentinel-5P and GEOS Forward Processing datasets (GEOS-FP) to develop a new approach for the daily estimation of full-coverage 5 km (0.05°) ambient concentrations of $PM_{2.5}$ and PM_{10} over China. The estimation function is obtained by fusing the multisource (Sentinel-5P TROPOMI, GEOS Forward Processing, and ground-based stations) data via ensemble machine learning methods, such as the light gradient boosting machine. On the other hand, Han et al. (2022) did an interpolation-based fusion of Sentinel-5P TROPOMI, elevation, and regulatory grade ground station data for producing spatially continuous maps of $PM_{2.5}$ concentrations over Thailand using different machine learning algorithms and comparing their accuracy. Furthermore, in their study, Son et al. (2022) proposed a deep learning-based surface $PM_{2.5}$ estimation method using the attentive interpretable tabular learning neural network (TabNet) with five Sentinel-5P TROPOMI products (NO_2, SO_2, O_3, CO, HCHO) over Thailand. They have tested the capability to estimate $PM_{2.5}$ without aerosol optical property, which was used more traditionally. Moreover, they highlighted CO as the most influential chemical component and related it to the seasonal burning in southeast Asia. In contrast to traditional studies, Li et al. (2022) proposed a knowledge-informed neural network model for joint estimation of $PM_{2.5}$ and O_3 over China, in which satellite observations, reanalysis data, and ground station measurements are used. Their conclusion was that the joint estimation model achieves performance comparable to that of the separate estimation model but with higher efficiency. Ahmed et al. (2022) developed a convolutional neural network (CNN) model which uses Sentinel-5P TROPOMI data of seven different pollutants (AI, CH_4, CO, HCHO, NO_2, O_3, SO_2), as auxiliary variables to estimate daily average concentrations of $PM_{2.5}$ in various cities in Pakistan from May 2019 to April 2020.

This research follows a new approach developed by Mamić et al. (2022) for accurately estimating atmospheric concentrations of $PM_{2.5}$ and PM_{10} using Sentinel-5P and other open-source remote sensing data from the Google Earth Engine (GEE) platform, a geospatial processing platform created to store and analyze enormous data sets for analysis and decision making. As noted by Gorelick et al. (2017), GEE's data catalog contains a repository of publicly available geospatial datasets, including observations from various satellite and aerial imaging systems in optical and non-optical wavelengths,

environmental variables, weather and climate forecasts and hindcasts, land cover, topographic, and socioeconomic datasets. All of this data is pre-processed into a ready-to-use, but the information-preserving format allows efficient access and removes numerous obstacles associated with data management. In this study, machine learning methods have been used to create models that can effectively determine air quality. Machine learning approaches have been recognized as useful and accurate in developing spatially explicit models. Random forest was employed in this study, as one of the most commonly used algorithms when it comes to estimating $PM_{2.5}$ and PM_{10} at both, national (Chen et al., 2018a, b; Stafoggia et al., 2019; Shao et al., 2020; Zhao et al., 2020) and regional (Huang et al., 2018; Yang et al., 2020) level.

Furthermore, several studies (Cichowicz et al., 2017; Xiao et al., 2015) have shown significant differences in the dispersion of atmospheric air pollution between heating and non-heating seasons. Accordingly, the main purpose of this study is to develop models that can be efficiently used to estimate the $PM_{2.5}$ and PM_{10} in the Republic of Croatia for non-heating (June, July, and August 2021) and heating (December 2021, January, and February 2022) season at a national and regional scale. The motivation to develop regional-scale models lies in the variability mentioned above of PM. Therefore, regional models will try to tackle the questions of composition and concentration of $PM_{2.5}$ and PM_{10} between different climatic and geographical features. The performance of developed models is evaluated using the metrics of the mean absolute error (MAE), the root mean squared error (RMSE), and the Pearson correlation coefficient (r). In addition, in situ and estimated $PM_{2.5}$ seasonal values were compared to those available by Copernicus Atmosphere Monitoring Service (CAMS) on the national level.

The research objectives can be divided into two main sections: (i) developing $PM_{2.5}$ and PM_{10} models for non-heating and heating season on a national scale and (ii) developing $PM_{2.5}$ and PM_{10} models for non-heating and heating season on a regional scale.

The structure of this paper is as follows. The second section introduces the study area, describes the used materials, and provides an overview of the methodology. The third section reveals the results of this research, followed by a discussion. In the final section, conclusions are provided.

Materials and methods

Study area

The Republic of Croatia has been chosen as the study area for this research (Fig. 1). Concerns about air pollution in Croatia increased rapidly last year after the Institute for Health Metrics and Evaluation (IHME) announced that Croatia ranks fifth in the European Union (EU) in terms of deaths caused by polluted air, putting the country at the bottom of the EU in terms of air quality (Index, 2021).

The Republic of Croatia is a country in Southeast Europe, covering 56,594 km², and its capital is Zagreb. Croatia is primarily a lowland country. The lowlands (terrain below 200 m absolute altitude) represent 53.4% of Croatia's territory, the hills (200 to 500 m absolute altitude) account for 25.6%, and the mountains and mountainous areas (above 500 m absolute altitude) account for 21.0%. The horseshoe shape reflects the importance of the continental and coastal regions, which are primarily linked by the karst mountain region (Hrvatska enciklopedija, 2021).

Biogeographical boundaries were gathered from the Emerald Network countries and EU member states. These were combined to create a map of biogeographical areas independent of political borders and cover all of Europe. The layer of biogeographical regions for all of Europe is available through the European Environment Agency (EEA, 2016). Accordingly, Croatia has three different biogeographical regions: Alpine, Continental, and Mediterranean.

The Alpine region of Croatia has a relatively cold and harsh climate, high altitudes, and a complex topography of the Dinaric Alps. With less than 10,000 residents, Ogulin and Gospić are the two largest towns in this region, the least inhabited in all of Croatia.

The Continental region of Croatia has a relatively flat landscape and a climate of strong contrasts between cold winters and hot summers. The Pannonian plain mainly influences it. The most populated cities in this region are Zagreb, Osijek, Slavonski Brod, Karlovac, and Varaždin.

The climate of Croatia's Mediterranean region is known for having hot, dry summers and humid, and chilly winters. However, it can also be notoriously unpredictable, with sudden rainstorms or bursts of
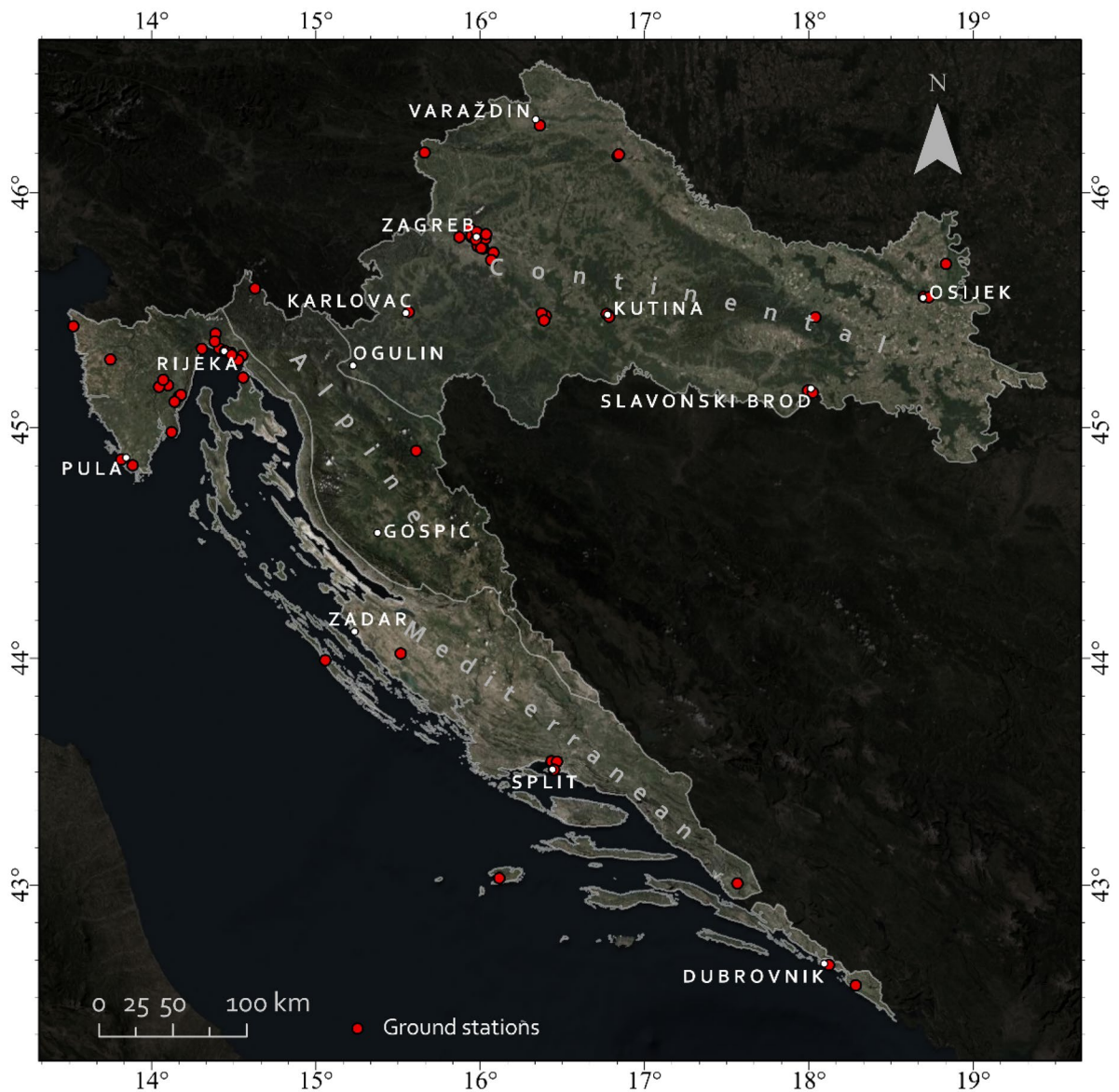
Fig. 1 Map of the Republic of Croatia with biogeographical regions and ground stations of the Croatian National Network for Continuous Air Quality Monitoring

strong winds occurring at different times of the year. The Adriatic Sea greatly influences it. Split, Rijeka, Zadar, and Pula are the largest cities in this region.

Materials

Since the composition of $PM_{2.5}$ and PM_{10} varies on multiple geographical and meteorological factors, we use multiple remote sensing data freely available on the GEE platform in this study. On the other hand, in situ data from the Croatian National Network for Continuous Air Quality monitoring stations will be used for validation as ground truth data.

In situ data

The ground stations in this study measure $PM_{2.5}$ and PM_{10} automatically every hour in μg/m^3 units. The

collected data are freely available on the official website of the Croatian National Network for Continuous Air Quality Monitoring (ISZZ, 2022). First, the $PM_{2.5}$ and PM_{10} raw hourly data were downloaded for all available stations for non-heating (June, July, and August 2021) and heating (December 2021 and January and February 2022) seasons in Croatia.

In Fig. 1, it is noticeable that the ground air quality monitoring stations are quite uniformly distributed at the national level. However, a closer look into their regional spatial distribution reveals an insufficient number of ground stations in the Alpine region, where data from just two stations were available for modeling, the non-heating season. On the other hand, only one station is available to model the Alpine region's heating season. That being said, the lack of stations would significantly disrupt the stability and the validity of the models; therefore, their creation for the Alpine region was abandoned. Regarding the Continental region, ground stations are evenly distributed throughout the region, and data from eight stations were used for $PM_{2.5}$ estimation in the non-heating season and from 11 for PM_{10}. In the heating season, nine stations were available for $PM_{2.5}$ and 14 for PM_{10}. The irregular shape of the Croatian Mediterranean region makes it geographically challenging for modeling. In addition, there is a lack of ground stations in the central and southern parts of the region, especially on islands.

Nonetheless, data from 10 stations were available for $PM_{2.5}$ estimation in both seasons. For PM_{10}, 16 stations were available in the non-heating and 17 in the heating season. Despite everything, it was decided to develop $PM_{2.5}$ and PM_{10} models for the Mediterranean region. On the national scale, due to some missing or invalid in situ data in the observed time range, data from 20 ground stations were used to estimate $PM_{2.5}$ for both seasons, and for PM_{10}, 30 stations were used for non-heating and 32 for the heating season.

Remote sensing data

The main data used to estimate this research's $PM_{2.5}$ and PM_{10} values is pre-processed L3 Sentinel-5P TROPOMI data of AI, CO, HCHO, NO_2, O_3, and SO_2. On the other hand, meteorological data used in this study is from the National Oceanic and Atmospheric Administration (NOAA) which provides a dataset consisting of selected model outputs as gridded forecast variables through its Global Forecast System (GFS), which is updated four times daily (every 6 h). GFS data used are land surface temperature 2 m above the ground in °C (LST), specific humidity 2 m above ground in kg/kg (kilogram of water per kilogram of air) (HUM), and U and V component of wind 10 m above ground in m/s (U-WIND and V-WIND). Geographical data on elevation (DEM) used in this study is from NASA's Shuttle Radar Topography Mission (SRTM), and slope data was derived from it. Moreover, soil pH data at ground

Table 1 Remote sensing data from GEE used in this study (Mamić et al., 2022)

Parameter	Description	Source	Unit	Temporal resolution	Spatial resolution
AI	Aerosol index	TROPOMI	/	daily	1113.2 m
CO	Carbon monoxide	TROPOMI	mol/m²	daily	1113.2 m
HCHO	Formaldehyde	TROPOMI	mol/m²	daily	1113.2 m
NO_2	Nitrogen dioxide	TROPOMI	mol/m²	daily	1113.2 m
O_3	Ozone	TROPOMI	mol/m²	daily	1113.2 m
SO_2	Sulphur dioxide	TROPOMI	mol/m²	daily	1113.2 m
LST	Land surface temperature	NOAA	°C	6 h	27 890 m
HUM	Specific humidity	NOAA	kg/kg	6 h	27 890 m
U-WIND	Eastward wind	NOAA	m/s	6 h	27 890 m
V-WIND	Northward wind	NOAA	m/s	6 h	27 890 m
DEM	Elevation	NASA SRTM	m	/	30 m
SLOPE	Slope	NASA SRTM	°	/	30 m
SOIL_pH	Soil pH	Hengl, 2018	pH*10	/	250 m

level was retrieved from the map made by Hengl in 2018 available on GEE. Besides, information on all used datasets from GEE is given in Table 1.

Methodology

The approach to estimate $PM_{2.5}$ and PM_{10} from multiple remote sensing data used in this study is shown in Fig. 2. All data collected were pre-processed for missing or invalid data before being matched spatially and temporally. The next step was to create new parameters from the main and auxiliary data to improve the stability of future models. Finally, all data were imported into Weka software for the attribute selection process and modeling by random forest

algorithm. Furthermore, accuracy assessment was done using r, MAE, and RMSE. Besides, GIS software was used to create spatial distribution maps of estimated $PM_{2.5}$ and PM_{10}.

Data pre-processing

Since the different spatial resolutions within the used datasets, only the pixel of the exact location of each ground station was extracted using the vector layer of ground stations imported in GEE. Regarding TROPOMI data, SO_2 was missing for the winter months (December 2021 and January and February 2022), so it was excluded from heating season models. Furthermore, there are also some missing CO, HCHO, and NO_2 data for the winter months due to the greater

Fig. 2 The flowchart of the approach to estimate $PM_{2.5}$ and PM_{10} followed by this study
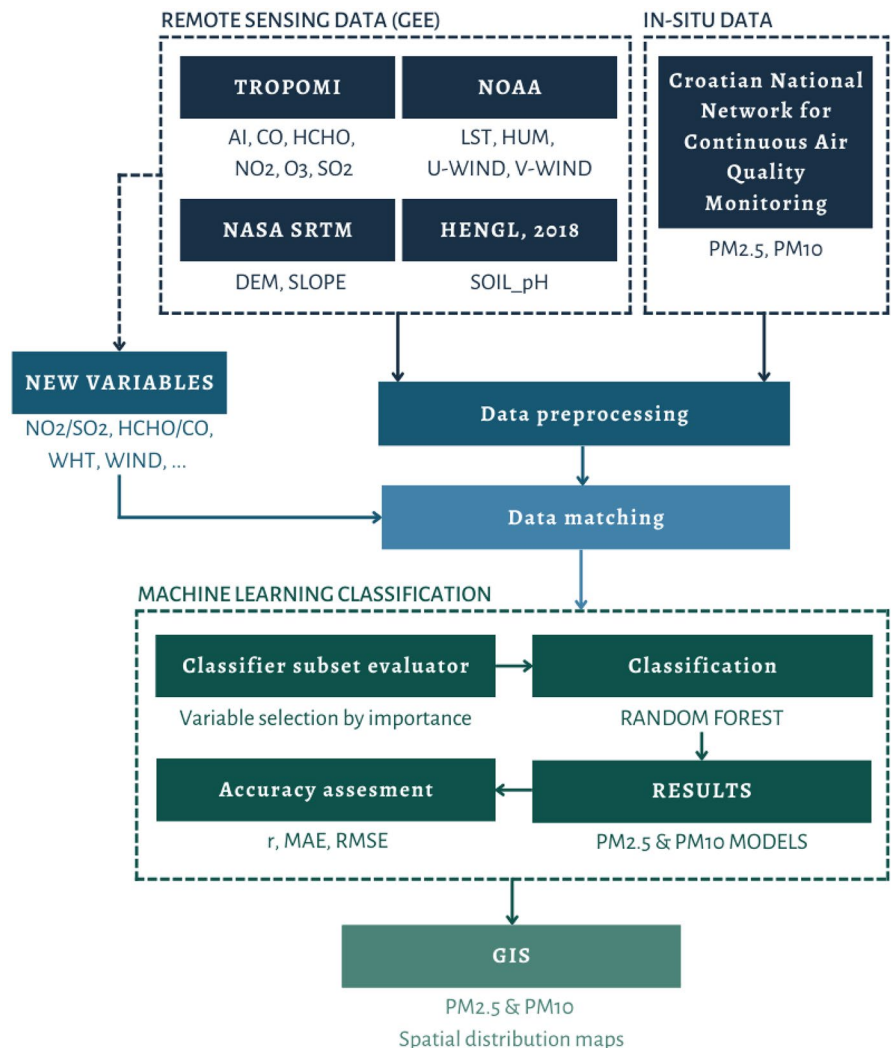
amount of cloudy days. Moreover, before July 2021, there was reported an error in the system for measuring AI so AI data from June 2021 was excluded from modeling, too. On the other hand, there were some missing and invalid (negative) in situ data, so those rows were removed from the dataset. Also, too high or too low ground measurements which differed too much from the rest of the dataset were deleted.

Data matching

Variables were matched spatially concerning each ground station's location and temporally regarding each dataset's temporal resolution. The Croatian National Network for Continuous Air Quality Monitoring Automatic ground stations measure atmospheric $PM_{2.5}$ and PM_{10} every hour. On the other hand, TROPOMI captures daily images above Croatia between 10:00 AM and 1:00 PM CET. On that basis, TROPOMI data was matched with the in situ data of the exact hour. Furthermore, NOAA's meteorological data are updated every 6 h (12:00 AM, 6:00 AM, 12:00 PM, and 6:00 PM) and were matched to the rest of the dataset based on the data for 12:00 PM. Geographical variables are temporally independent, so they were matched based only on location.

Covariates

The original parameters of Table 1 were expanded with the new ones (Table 2) based on the similarities they share to improve the models' stability and to find out the composition and therefore possible sources of $PM_{2.5}$ and PM_{10} in the observed period and between different biogeographical regions.

Modeling $PM_{2.5}$ and PM_{10}

This study employed a random forest algorithm to estimate $PM_{2.5}$ and PM_{10} values from multiple remote sensing data. The random forest algorithm was first proposed by Breiman (2001), and since that, it has been extremely successful as a general-purpose classification and regression method. The method, which combines several randomized decision trees and averages their predictions, has shown excellent performance in settings with a large number of variables compared to the number of observations. Furthermore, it is versatile enough to be applied to large-scale

Table 2 Covariates created from remote sensing data (Mamić et al., 2022)

No	Parameter
1	$(NO_2 + SO_2)/(NO_2\text{-}SO_2)$
2	NO_2/SO_2
3	$HCHO/CO$
4	$(CO + HCHO)/CO\text{-}HCHO)$
5	$O_3/(NO_2 + SO_2 + CO)$
6	$AI*(NO_2/SO_2)$
7	$O_3/(NO_2/SO_2)$
8	$O_3/((NO_2 + SO_2)/(NO_2\text{-}SO_2))$
9	$SQRT(1/(NO_2 + SO_2 + O_3))$
10	$(AI + HUM)/(AI\text{-}HUM)$
11	$(AI + DEM)/(AI\text{-}DEM)$
12	$(CO + NO_2)/CO\text{-}NO_2)$
13	$(CO + O_3)/(CO\text{-}O_3)$
14	$WIND$[a]
15	WHT[b]
16	$(WHT + AI)/(WHT\text{-}AI)$
17	$(WHT + O_3)/(WHT\text{-}O_3)$
18	$(CO + SO_2)/(CO\text{-}SO_2)$
19	$S5$[c]
20	$(S5 + WHT)/(S5\text{-}WHT)$
21	$(S5 + WIND)/(S5\text{-}WIND)$

[a]$(U\text{-}WIND + V\text{-}WIND)/2$

[b]$(((U\text{-}WIND + V\text{-}WIND)/2) + HUM + LST)/3$

[c]$(AI + CO + HCHO + NO_2 + O_3 + SO_2)/6$

problems, adaptable to a variety of ad hoc learning tasks, and returns information on each variable importance (Biau & Scornet, 2016).

For the analyses, we used Weka (Waikato Environment for Knowledge Analysis), open-source software—first for the process of attribute selection and finally, for classification by random forest algorithm. Weka is a large collection of Java class libraries that implement a wide range of state-of-the-art machine learning and data mining algorithms (Witten et al., 1999). However, as mentioned above, prior to modeling, it was necessary to choose only the best parameters for each model. For that purpose, the Weka Classifier Subset Evaluator tool was used for the random forest algorithm with a percentage split of 70 to find the most important variables. Starting with the best-performing parameter, the Classifier Subset Evaluator progressively adds parameters to the model, one at a time, in order of their ranking. At

each step, the accuracy of the classifier is measured on the test set using only the current subset of parameters. This process continues until the accuracy no longer improves. The Classifier Subset Evaluator is a useful way to reduce the dimensionality of the data while maintaining or even improving classification performance. By selecting only the most informative parameters, the method can improve the efficiency of machine learning algorithms and reduce overfitting. All remaining attributes were removed once the best had been identified, and then, a random forest classifier was used with data split into training and testing portions of 70% and 30%, respectively. Once the model was developed, it was saved, and all models in this study were trained separately by the same procedure described above. It is important to note that each model should use at least one parameter directly related to air pollution. The total number of instances and parameters to build each model is shown in Table 3.

Accuracy assessment

The accuracy of the machine learning models may be validated using various metrics. However, the mean absolute error (MAE) and the root mean squared error (RMSE) are commonly used as evaluation metrics when estimating air pollutants (Hu et al., 2017; Larkin et al., 2017; Ayturan et al., 2018; Rybarczyk & Zalakeviciute, 2018; Van Roode et al., 2019; Masih, 2019).

MAE refers to the mean of the absolute error values of each prediction for all instances of the test dataset. The prediction error represents the difference between that case's actual and predicted value. MAE treats all individual differences with equal weight.

$$MAE = \sqrt{\sum_{i=1}^{n} \frac{|y_i - \widehat{y}_i|}{n}} \tag{1}$$

The other metric used in this study is RMSE, as a standard way to measure the error of a model in predicting quantitative data. It's the square root of the average of squared differences between predicted and actual values.

$$RMSE = \sqrt{\sum_{i=1}^{n} \frac{(y_i - \widehat{y}_i)^2}{n}} \tag{2}$$

In given equations $\widehat{y}_1, \widehat{y}_2, \ldots, \widehat{y}_n$ represent predicted values, $y_1, y_2, \ldots, y_n$ are actual values, and n is the number of instances.

Results and discussions

Used variables

As said earlier, the composition of $PM_{2.5}$ and PM_{10} varies due to the various climatic and geographical features. Also, between non-heating and heating seasons. Therefore, including multiple parameters in their

Table 3 Number of instances and parameters used to create each model

Model	Pollutant	Number of instances	Number of parameters
Continental: non-heating season	$PM_{2.5}$	274	11
Continental: non-heating season	PM_{10}	379	6
Mediterranean: non-heating season	$PM_{2.5}$	281	10
Mediterranean: non-heating season	PM_{10}	397	8
Continental: heating season	$PM_{2.5}$	322	10
Continental: heating season	PM_{10}	479	7
Mediterranean: heating season	$PM_{2.5}$	355	7
Mediterranean: heating season	PM_{10}	617	12
Croatia: non-heating season	$PM_{2.5}$	602	10
Croatia: non-heating season	PM_{10}	823	8
Croatia: heating season	$PM_{2.5}$	690	10
Croatia: heating season	PM_{10}	1109	10

Table 4 Parameters used to develop $PM_{2.5}$ and PM_{10} seasonal models in the Continental region of Croatia

Pollutant	Season	Parameters
$PM_{2.5}$	Non-heating	NO_2, LST, HUM, DEM, NO_2/SO_2, (CO+HCHO)/CO-HCHO), AI*(NO_2/SO_2), O_3/$((NO_2+SO_2)/(NO_2\text{-}SO_2))$, (CO+$NO_2$)/CO-$NO_2$), WIND, (WHT[a]+$O_3$)/(WHT-$O_3$)
$PM_{2.5}$	Heating	AI, CO, O_3, V-WIND, SOIL_pH, DEM, (CO+HCHO)/(CO-HCHO), WIND, (WHT+AI)/(WHT-AI), (WHT+O_3)/(WHT-O_3)
PM_{10}	Non-heating	LST, SOIL_pH, SLOPE, (CO+HCHO)/(CO-HCHO), AI*(NO_2/SO_2), (WHT+O_3)/(WHT-O_3)
PM_{10}	Heating	AI, LST, HUM, SOIL_pH, SLOPE, (AI+HUM)/(AI-HUM), (AI+DEM)/(AI-DEM)

[a](((U-WIND+V-WIND)/2)+HUM+LST)/3

estimation is crucial to improve the models' stability and accuracy and discover the composition of specified air pollutants and their possible sources. The selected parameters used to develop models and estimate $PM_{2.5}$ and PM_{10} seasonal values on the national and regional scale of Croatia are shown in Tables 4, 5, and 6.

Looking at the attributes used to develop seasonal models of $PM_{2.5}$ and PM_{10} for the Continental region of Croatia, it can be noticed that models use from six to 11 parameters. Furthermore, it is also noticeable how the $PM_{2.5}$ models use more parameters than the PM_{10} models. Both $PM_{2.5}$ models share four parameters: DEM, (CO+HCHO)/(CO-HCHO), WIND, and (WHT+O_3)/(WHT-O_3), while PM_{10} models share LST, SOIL_pH, and SLOPE parameters.

When looking at the attributes used to build seasonal models of $PM_{2.5}$ and PM_{10} for Croatia's Mediterranean region, it is noticeable that all models use from seven to 12 parameters, similar to those of the Continental region. Both $PM_{2.5}$ models have three parameters in common: U-WIND, DEM, and (WHT+AI)/(WHT-AI), whereas PM_{10} models have five: CO, NO_2, V-WIND, SLOPE, and (AI+DEM)/(AI+DEM) (AI-DEM). Since the Mediterranean region of Croatia is known for having strong winds, all four models incorporate at least one wind component parameter.

All national models used a similar number of attributes (8 to 10). For both $PM_{2.5}$ models, CO, LST, U-WIND, and SLOPE parameters are used. On the other hand, both PM_{10} models share following parameters: NO_2, SOIL_pH, DEM, SLOPE (AI+HUM)/(AI-HUM), and (WHT+AI)/(WHT-AI). Moreover, in all four models, the SLOPE parameter is used.

As previously noted, several studies to date have employed TROPOMI data to estimate atmospheric $PM_{2.5}$ and PM_{10} concentrations. Therefore, to estimate $PM_{2.5}$ and PM_{10} over China Wang et al. (2021) used numerous parameters (30) from multiple sources, such as TROPOMI (SO_2, NO_2, and O_3), GEOS-FP (black carbon, organic carbon, nitrate, SO_4, dust, ammonium, sea salt, humidity, air temperature, U-WIND, V-WIND, total precipitable water vapor, Pbl top pressure, surface pressure, planetary boundary layer height, air density at surface, surface velocity scale, and evaporation from turbulence), MODIS (NDVI, fractions of forest, savanna, grassland, cropland, urban, and arid land), Open Street Map (road density), and GPW (population density). Feature importance analysis done by Wang et al. (2021) showed that for $PM_{2.5}$ estimation five the most significant variables are NO_2, U-WIND, V-WIND, SO_2, and Pbl top pressure. On the other hand, their analysis showed that for

Table 5 Parameters used to develop $PM_{2.5}$ and PM_{10} seasonal models in the Mediterranean region of Croatia

Pollutant	Season	Parameters
$PM_{2.5}$	Non-heating	SO_2, HUM, U-WIND, SOIL_pH, DEM, SLOPE, SQRT($1/(NO_2+SO_2+O_3)$), (CO+NO_2)/CO-NO_2), (CO+O_3)/(CO-O_3), (WHT[a]+AI)/(WHT-AI)
$PM_{2.5}$	Heating	NO_2, O_3, U-WIND, V-WIND, DEM, (WHT+AI)/(WHT-AI), (WHT+O_3)/(WHT-O_3)
PM_{10}	Non-heating	CO, NO_2, V-WIND, SLOPE, SQRT($1/(NO_2+SO_2+O_3$), (AI+DEM)/(AI-DEM), WIND, (CO+SO_2)/(CO-SO_2)
PM_{10}	Heating	AI, CO, NO_2, HUM, U-WIND, V-WIND, DEM, SLOPE, (AI+HUM)/(AI-HUM), (AI+DEM)/(AI-DEM), WHT, (WHT+O_3)/(WHT-O_3)

[a](((U-WIND+V-WIND)/2)+HUM+LST)/3

Table 6 Parameters used to develop $PM_{2.5}$ and PM_{10} seasonal models in Croatia

Pollutant	Season	Parameters
$PM_{2.5}$	Non-heating	CO, HCHO, LST, U-WIND, V-WIND, SOIL_pH, DEM, SLOPE, AI*(NO_2/SO_2), SQRT(1/ ($NO_2 + SO_2 + O_3$))
$PM_{2.5}$	Heating	CO, O_3, HUM, LST, U-WIND, SLOPE, HCHO/CO, (CO + NO_2)/(CO-NO_2), (CO + O_3)/(CO-O_3), WHT[a]
PM_{10}	Non-heating	NO_2, SOIL_pH, DEM, SLOPE, ($NO_2 + SO_2$)/(NO_2-SO_2), AI*(NO_2/SO_2), (AI + HUM)/(AI-HUM), (WHT + AI)/(WHT-AI)
PM_{10}	Heating	HCHO, NO_2, HUM, U-WIND, V-WIND, SOIL_pH, DEM, SLOPE, (AI + HUM)/(AI-HUM), (WHT + AI)/ (WHT-AI)

[a](((U-WIND + V-WIND)/2) + HUM + LST)/3

PM_{10} estimation five of the most significant variables are NO_2, U-WIND, V-WIND, dust, and SO_2. When it comes to the study by Han et al. (2022) who were estimating $PM_{2.5}$ ambient concentrations for Thailand, they used only TROPOMI (AI, CO, HCHO, SO_2, NO_2, and O_3) and NASA SRTM elevation data (DEM), and their feature importance analysis showed NO_2 and SO_2 as the most unsignificant parameters with both long-term (1 month) and short-term (10 days) dataset. On the other hand, CO and AI were chosen as the most significant parameters regarding both datasets. A study done by Son et al. (2022) also focused on estimating the $PM_{2.5}$ concentrations in Thailand by combining multiple datasets: TROPOMI (CO, HCHO, SO_2, NO_2, and O_3), ERA5 (temperature – T, dew-point temperature – T_d, total evaporation, surface pressure, precipitation, U-WIND, V-WIND), ETOPO1 (DEM), and GlobCover (22 types of different land cover). Moreover, they approximated relative humidity and wind speed using Eqs. (3) and (4).

$$relhumidity = 100 * \frac{e^{\frac{17.625*T_d}{243.04*T_d}}}{e^{\frac{17.625*T}{243.04*T}}} \quad (3)$$

$$windspeed = \sqrt{U - WIND^2 + V - WIND^2} \quad (4)$$

Another study regarding $PM_{2.5}$ estimation using TROPOMI data was done by Li et al. (2022) as a joint estimation of $PM_{2.5}$ and O_3 over China, and they used TROPOMI (HCHO and NO_2), ERA5 (U-WIND, V-WIND, temperature, evaporation, total precipitation, surface pressure, surface, and top net solar radiation), CAMS (NO_2, HCHO, NO, $PM_{2.5}$, and O_3). As opposed to other mentioned studies, Ahmed et al. (2022) used only TROPOMI data (AI, CH_4, CO, HCHO, NO_2, O_3, SO_2) to estimate average concentrations of $PM_{2.5}$ in various cities in Pakistan.

Spatial distribution maps

Detecting $PM_{2.5}$ and PM_{10} hotspots is essential in finding possible sources of air pollution. Therefore, based on the developed models and in situ data, using ArcGIS Pro 2.8.3 software and minimum curvature spline technique the interpolation maps of $PM_{2.5}$ and PM_{10} were made on the national and regional scale of Croatia for both, non-heating and heating season (Figs. 3, 4, 5, 6, 7, and 8) to show the spatial distribution of the observed pollutants. The ground stations used for modeling are indicated as point data on the maps.

It is assumed that during the heating season, pollution will be higher. Furthermore, based on the geographical features of the Continental region, such as a relatively flat landscape and a climate of strong contrasts, it is assumed that it will be more polluted than the rest of the country.

The prediction maps are nearly identical to those created using in situ data. For the Continental region, the assumption that the heating season is more polluted was correct for $PM_{2.5}$, but for PM_{10}, the situation remained almost constant across seasons. On the other hand, for the non-heating season, $PM_{2.5}$ pollution is low, with some mild hotspots visible in the southeast part of the region, whereas for the heating season, pollution is high and focused on the region's west side. When it comes to PM_{10}, in both seasons, pollution is highest in the far east of the region, with one hotspot around the urban area of the city of Osijek. There are two other noticeable hotspots in the central part of the region around the urban area of the city of Zagreb and the town of Kutina.

The assumption that the heating season is more polluted in the Mediterranean region was incorrect; again, predicted values appear to be slightly lower than actual ones and slightly underestimate them. Moreover, in the Mediterranean region, pollution is generally low. A single-polluted coastal area in the northern part is likely influenced by the urban area of the city of Rijeka, which is located directly above it. For both pollutants, there is a decrease during the heating season.

The interpolation maps on a national scale give us visual insight between the ground and remote sensing data and show the seasonal variations of $PM_{2.5}$ and PM_{10} across the whole country. Prediction maps are nearly identical to those created using in-situ data. All conclusions drawn from regional maps are noticeable here. The Continental region stands out as the most polluted, with already mentioned hotspots located mostly in urban areas, and several studies (Fenger, 1999; Gulia et al., 2015; Kaplan et al.,
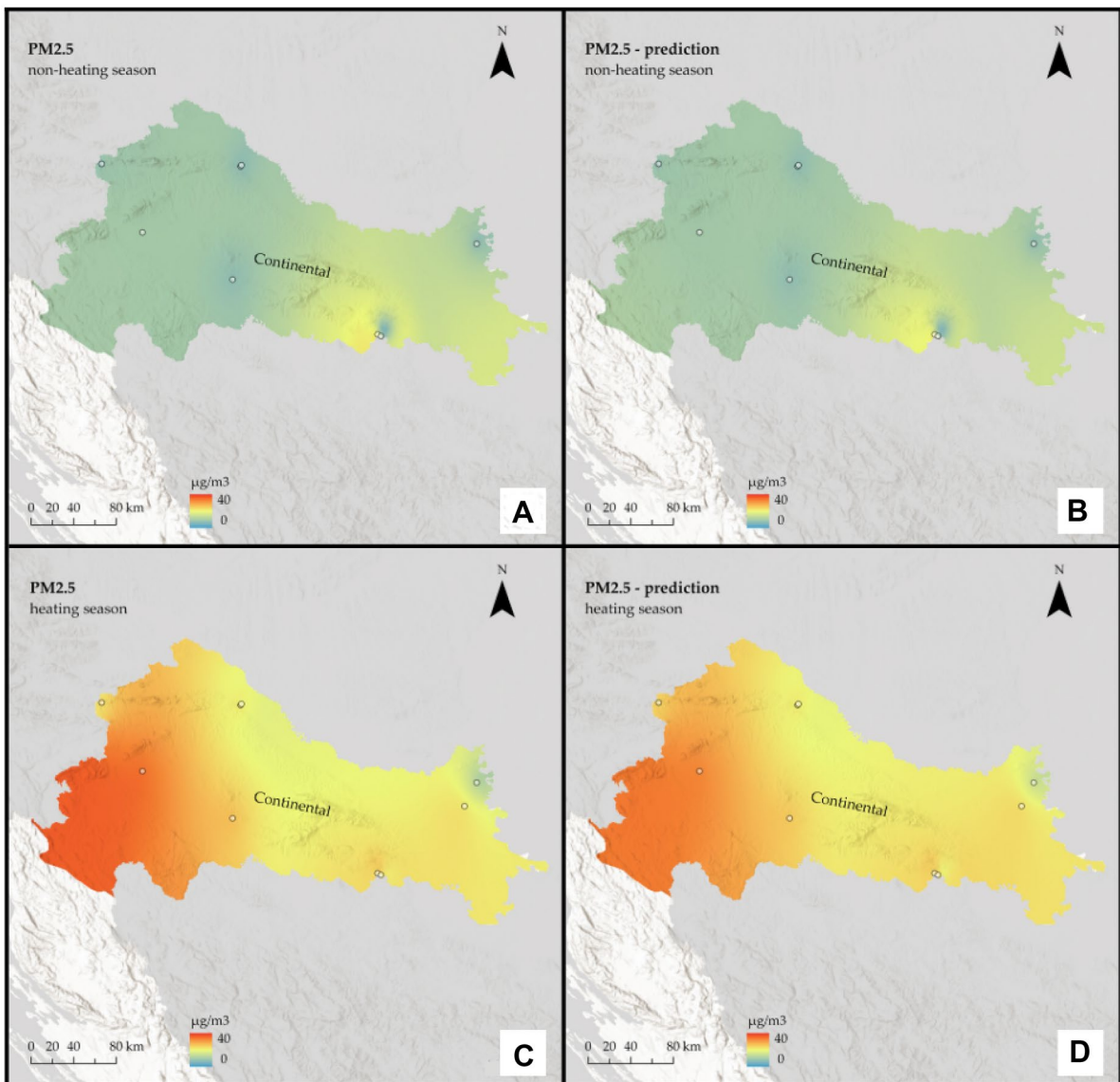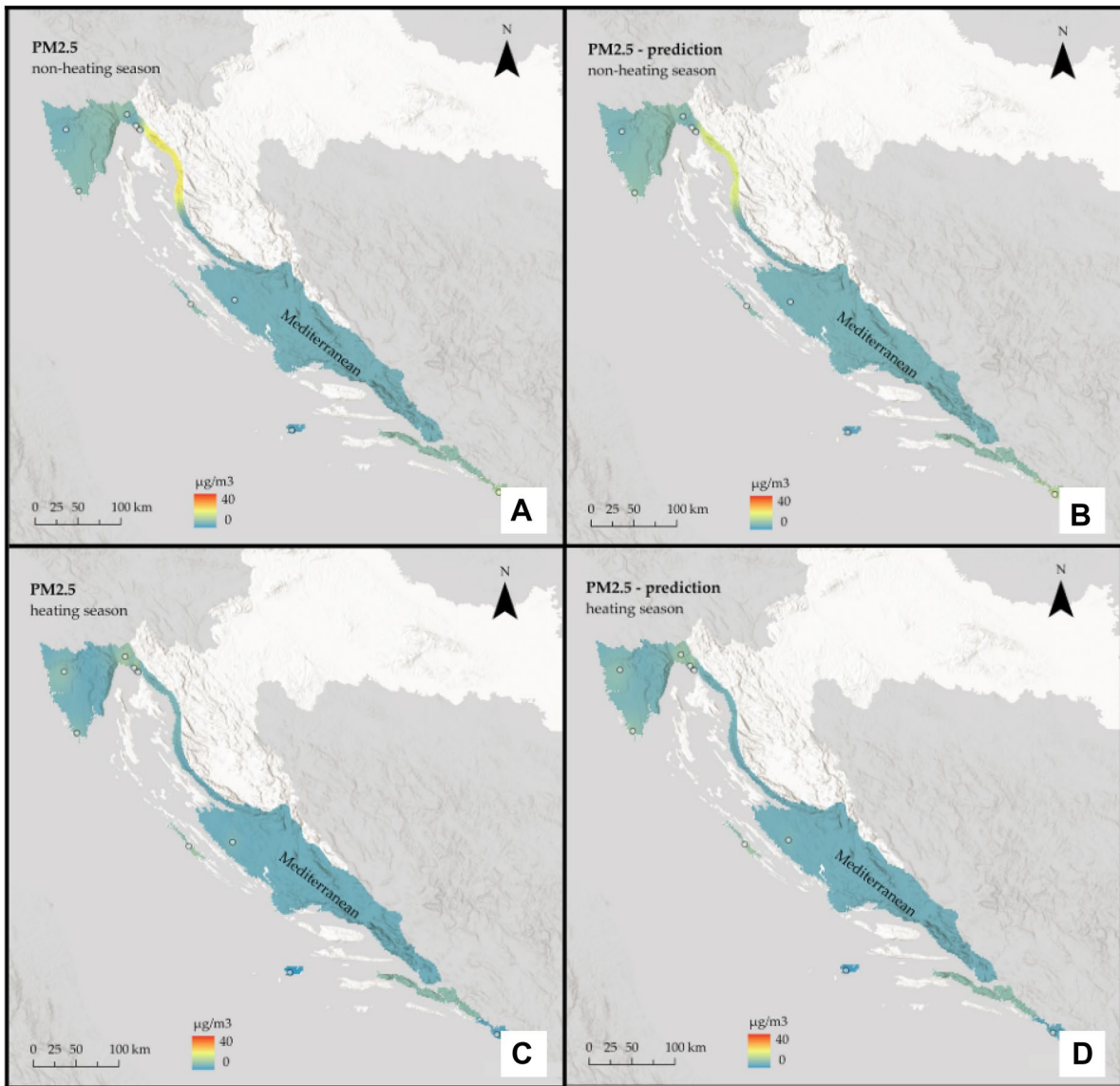

Fig. 3 Interpolated $PM_{2.5}$ concentrations for the Continental region of Croatia: **A** non-heating season from in situ data, **B** non-heating season from remote sensing data, **C** heating season from in situ data, and **D** heating season from remote sensing data
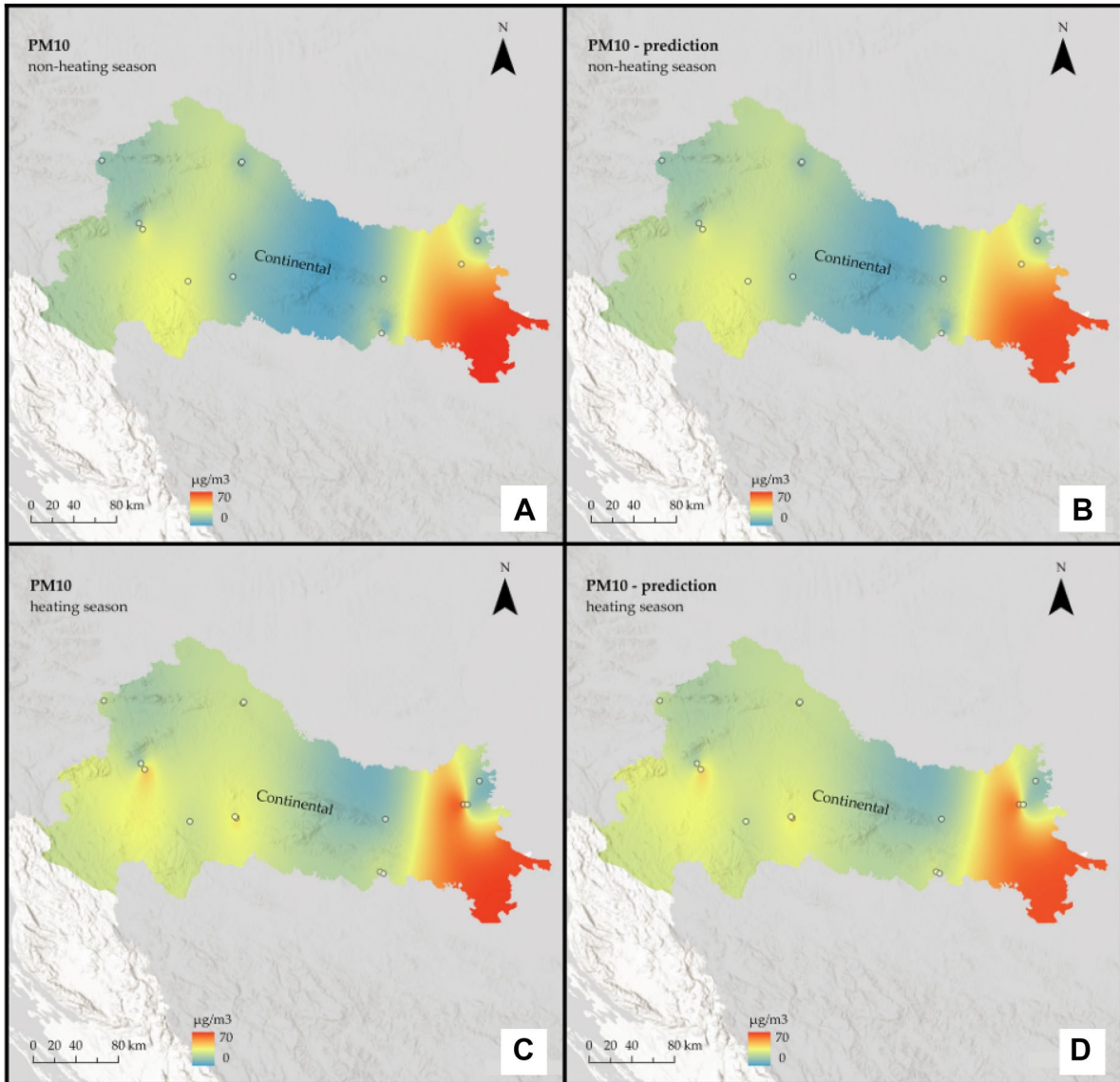
Fig. 4 Interpolated PM$_{10}$ concentrations for the Continental region of Croatia: **A** non-heating season from in situ data, **B** non-heating season from remote sensing data, **C** heating season from in situ data, and **D** heating season from remote sensing data

2019) suggest that cities and high population densities are causes of air pollution. The low pollution in Alpine and Mediterranean regions supports other studies that talk about the impact of altitude (Kaplan & Avdan, 2020; Mamić, 2021; Ning et al., 2018) and sea (Rosenfeld et al., 2002) on air quality. Furthermore, seasonal changes in PM$_{2.5}$ were also noticed in studies by Wang et al. (2021) and Li et al. (2022) who connected them with heating emissions and various meteorological conditions. Moreover, Wang et al. (2021) linked large PM$_{10}$ emissions with sand storms and dry weather, which cannot be applicable to our study area since neither of these conditions is typical for any part of Croatia.
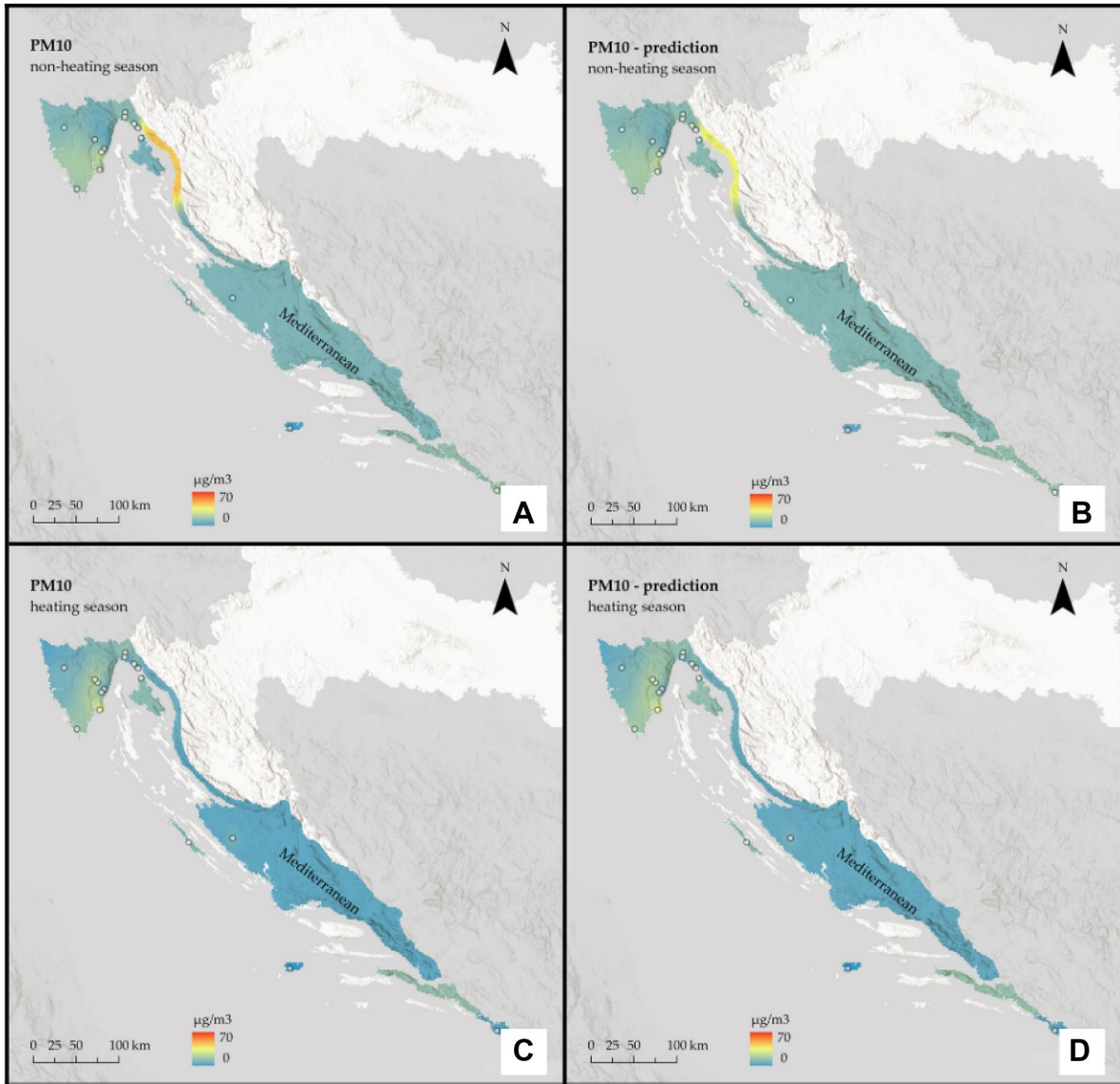
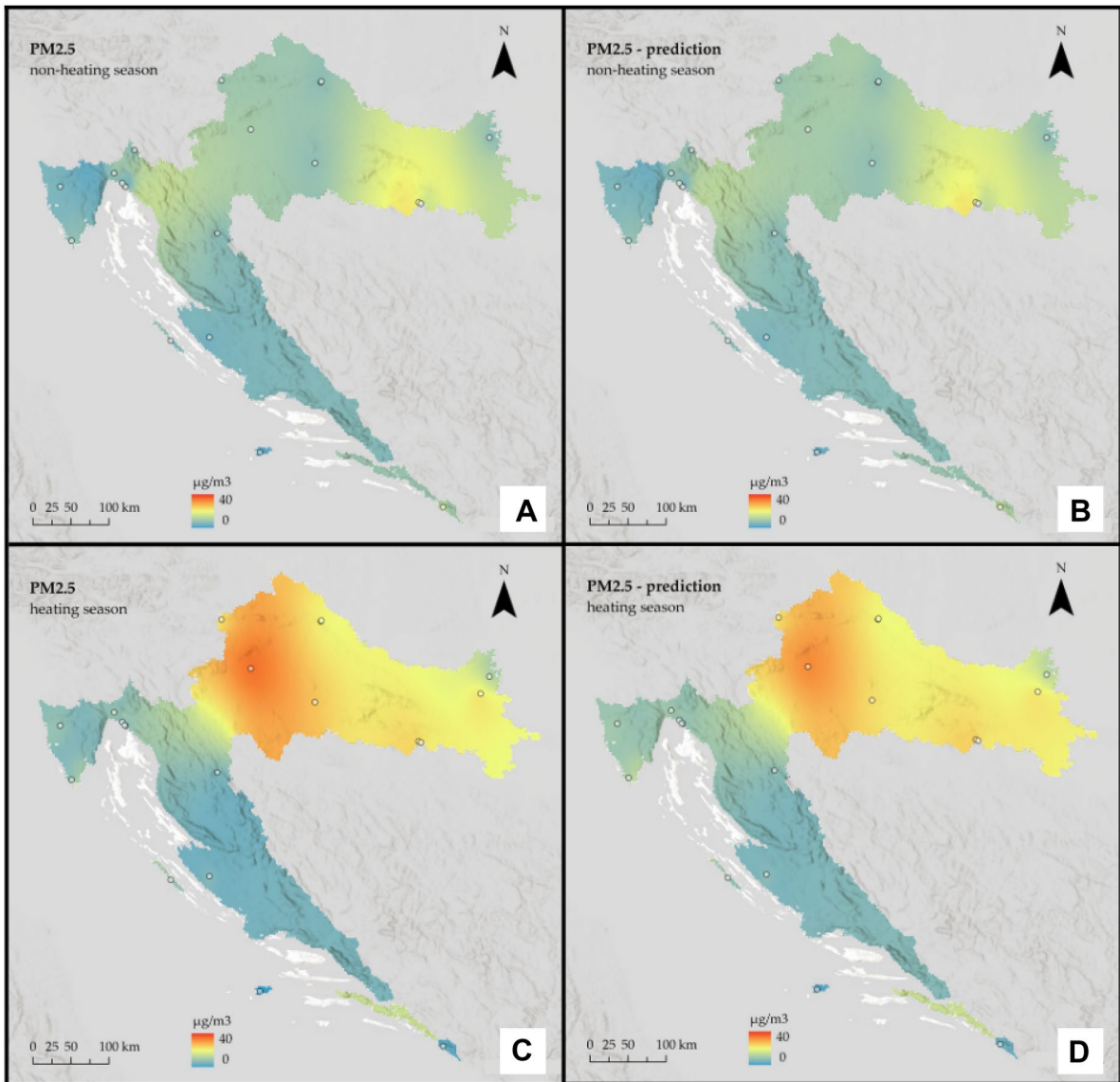**Fig. 5** Interpolated PM$_{2.5}$ concentrations for the Mediterranean region of Croatia: **A** non-heating season from in situ data, **B** non-heating season from remote sensing data, **C** heating season from in situ data, and **D** heating season from remote sensing data

Validation

The PM$_{2.5}$ and PM$_{10}$ models on the national and regional scale of Croatia for heating and non-heating seasons were developed and are shown in Figs. 9, 10 with their correlation coefficient (*r*), mean absolute error (MAE) and root mean squared error (RMSE).

The Continental region is regarded as the most polluted region in Croatia. All developed models have a high correlation that does not vary much with the seasons, making them highly stable. The accuracy of PM$_{2.5}$ models is higher than that of PM$_{10}$. On the other hand, the Mediterranean region is influenced by the sea and does not have high pollution.

Fig. 6 Interpolated PM$_{10}$ concentrations for the Mediterranean region of Croatia: **A** non-heating season from in situ data, **B** non-heating season from remote sensing data, **C** heating season from in situ data, and **D** heating season from remote sensing data

All developed models have a high correlation, which increases with the heating season. Similarly to models developed for the Continental region, the accuracy of PM$_{2.5}$ models is higher than that of PM$_{10}$.

The final objective of this research was to develop PM$_{2.5}$ and PM$_{10}$ models for the non-heating and heating season in Croatia using only open-source data from the GEE. The PM$_{2.5}$ models for both seasons show a high correlation with $r = 0.71$ (MAE = 3.23 μg/m^3 and RMSE = 5.27 μg/m^3) for non-heating and $r = 0.73$ (MAE = 6.30 μg/m^3 and RMSE = 8.82 μg/m^3) for the heating season. PM$_{10}$ models are a little less accurate, but still show moderate – $r = 0.59$ (MAE = 7.64 μg/m^3 and RMSE = 13.33 μg/m^3) for the non-heating season,
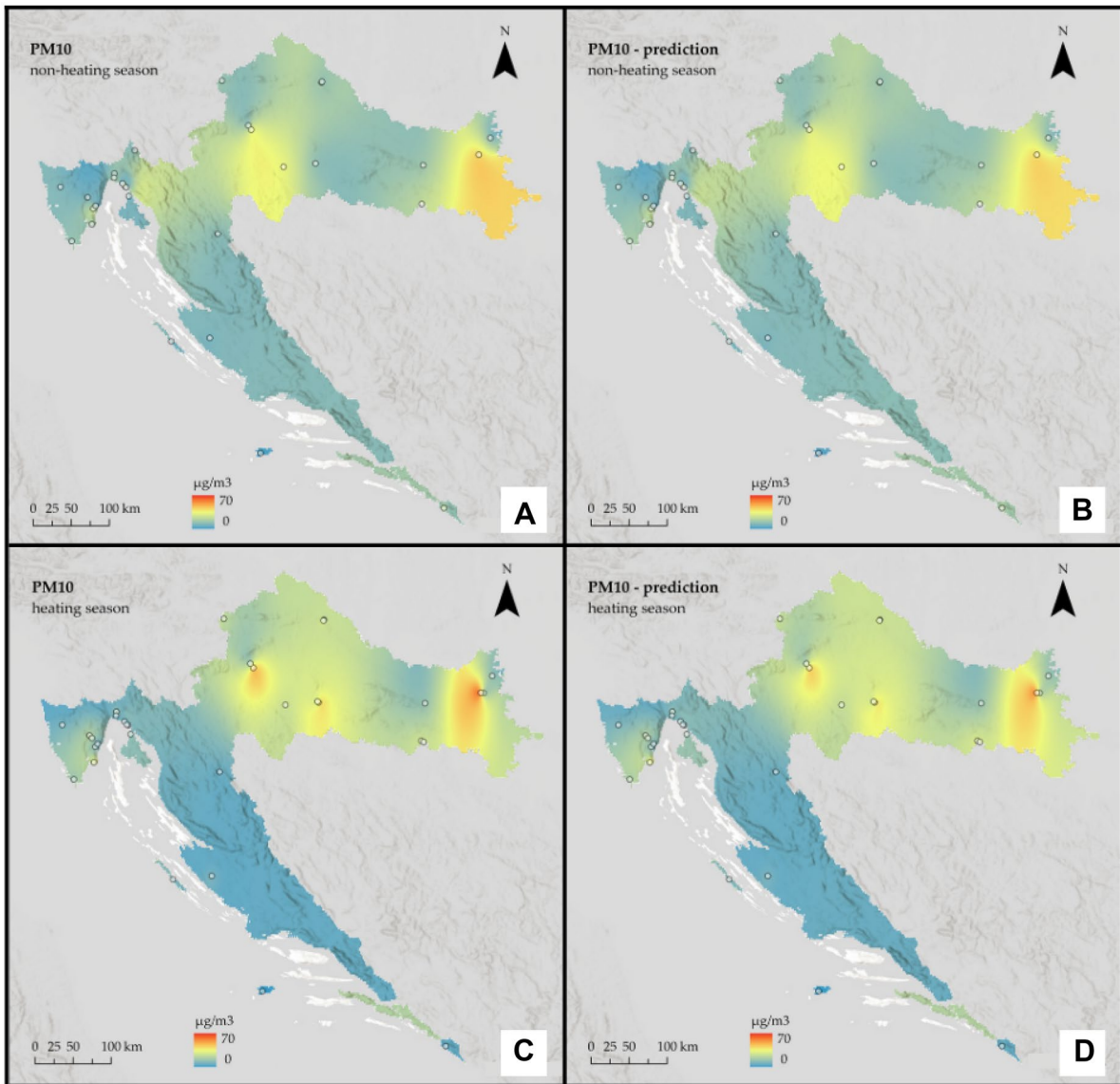
Fig. 7 Interpolated PM$_{2.5}$ concentrations for Croatia: **A** non-heating season from in situ data, **B** non-heating season from remote sensing data, **C** heating season from in situ data, and **D** heating season from remote sensing data

to high correlation – $r = 0.68$ (MAE $= 10.23$ μg/m^3 and RMSE $= 15.02$ μg/m^3) for the heating season.

To better understand the developed national models, we can look at one ground station and its values (Fig. 11). This is the most western station located on the Istrian peninsula close to the sea and is installed on an industrial waste management site. Here we can

see that the error between the actual and predicted values is the smallest for the PM$_{2.5}$ non-heating season model and is close to 0 for this station. For this station, the error is the highest for the PM$_{2.5}$ heating season model, even though this model has the highest accuracy among national models. The difference in seasons is clearly visible, where for the heating

Fig. 8 Interpolated PM_{10} concentrations for Croatia: **A** non-heating season from in situ data, **B** non-heating season from remote sensing data, **C** heating season from in situ data, and **D** heating season from remote sensing data

season we have higher values than in non-heating, as a justification for the seasonal temporal frame used in this research.

All $PM_{2.5}$ and PM_{10} models developed in this research are shown in Fig. 12, with their *r*, MAE, and RMSE. The *r* of each developed model is ranked from highest (green) to lowest (red). Low MAE and RMSE values are represented by red arrows, while yellow and blue arrows, respectively, represent medium and high values. Therefore, it is clear that all $PM_{2.5}$

models have shown better performance compared to PM_{10}. Moreover, almost all regional models outperformed national ones.

Comparison with CAMS data

The Copernicus Atmosphere Monitoring Service (CAMS) provides global and European data related to air pollution and health, greenhouse gas emissions, solar energy, and climate forcing. For the
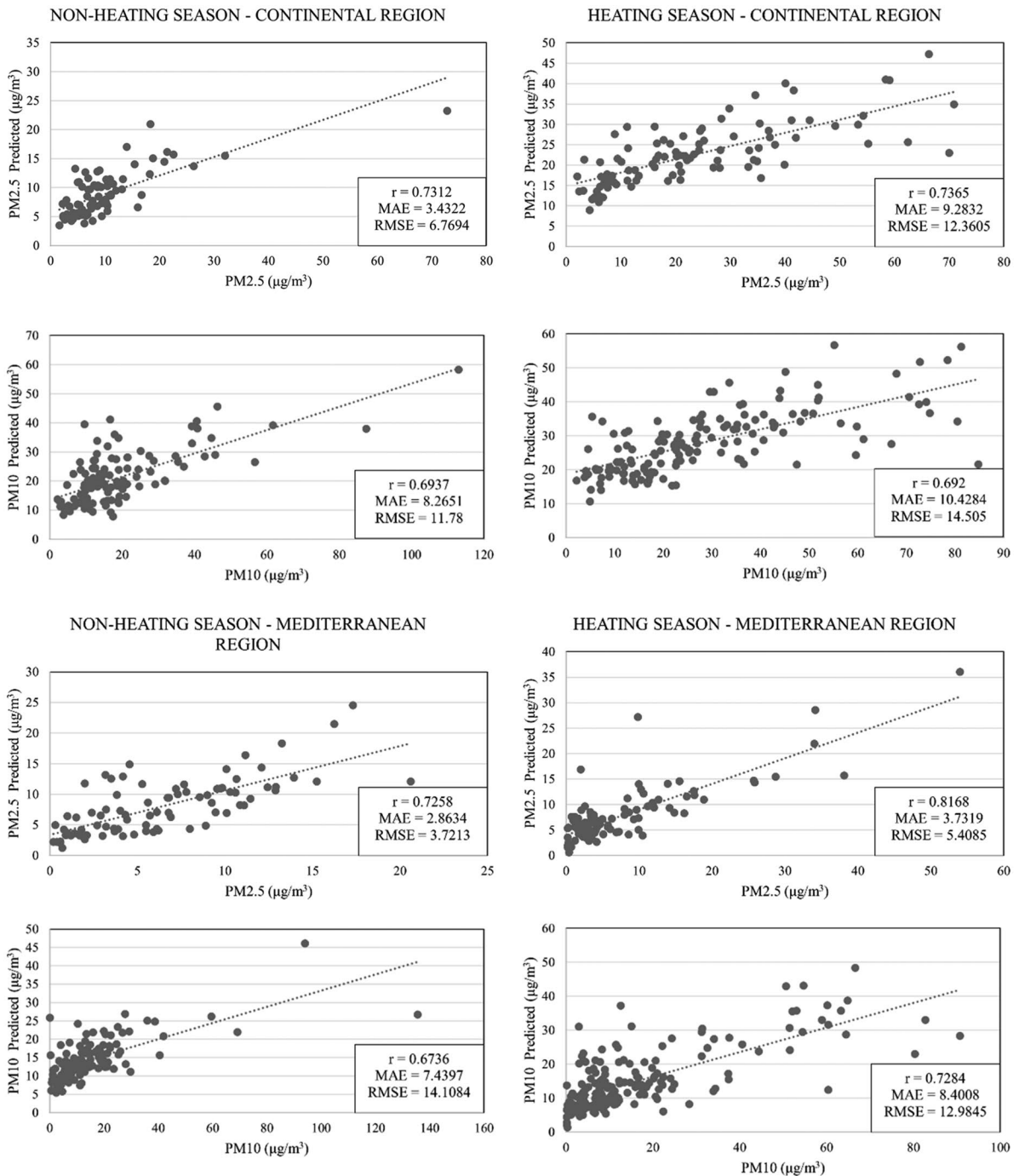
Fig. 9 Regional PM$_{2.5}$ and PM$_{10}$ prediction models

period observed by this research, available CAMS data is only that of PM$_{2.5}$ at the spatial resolution of 44 528 m. The averaged PM$_{2.5}$ data for the non-heating and heating season was collected for Croatia by GEE and compared with the in situ data and data predicted by developed models (Fig. 13).
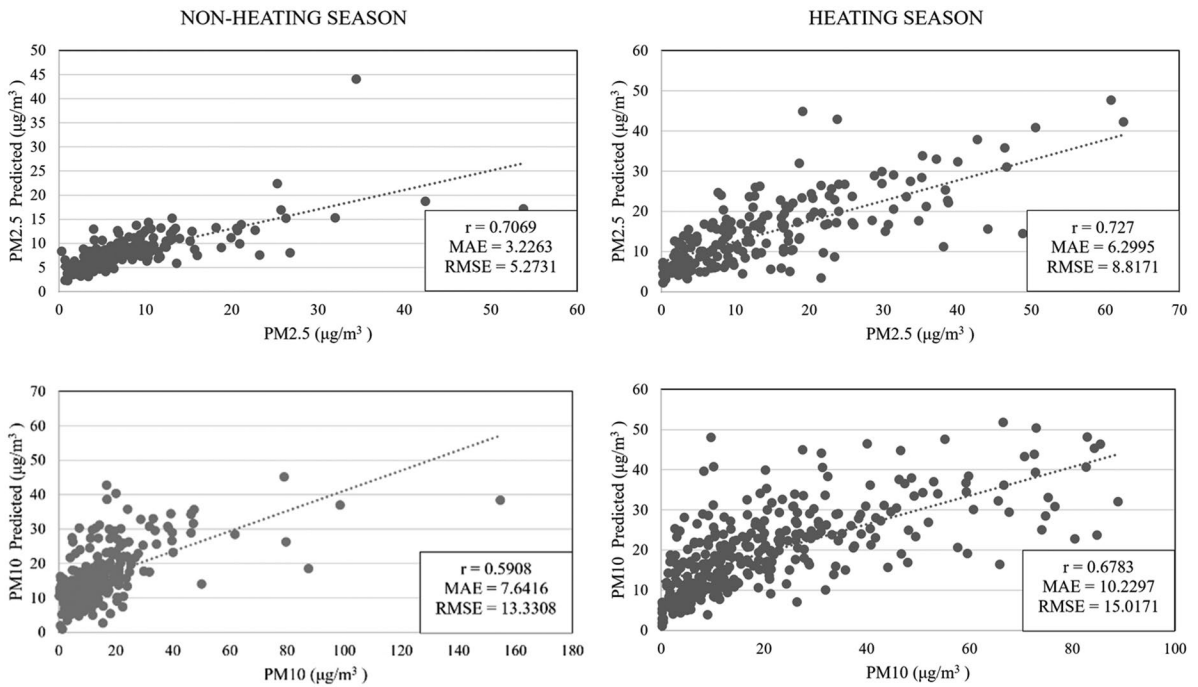
Fig. 10 National $PM_{2.5}$ and PM_{10} prediction models for Croatia

For the non-heating season, the CAMS data appear higher than most stations' actual values. On the other hand, the situation is the opposite for the heating season, where the CAMS data are underestimating the actual values. The CAMS data appear to be consistent throughout all stations and are unable to show sudden changes in the observed pollutant, which may be due to the lower spatial resolution of the data. All being said, CAMS is a valuable source of $PM_{2.5}$ data. However, on this scale, the models developed in this study proved to be a better solution.

Fig. 11 AMP Kaštijun ground station with actual and predicted data for all national models
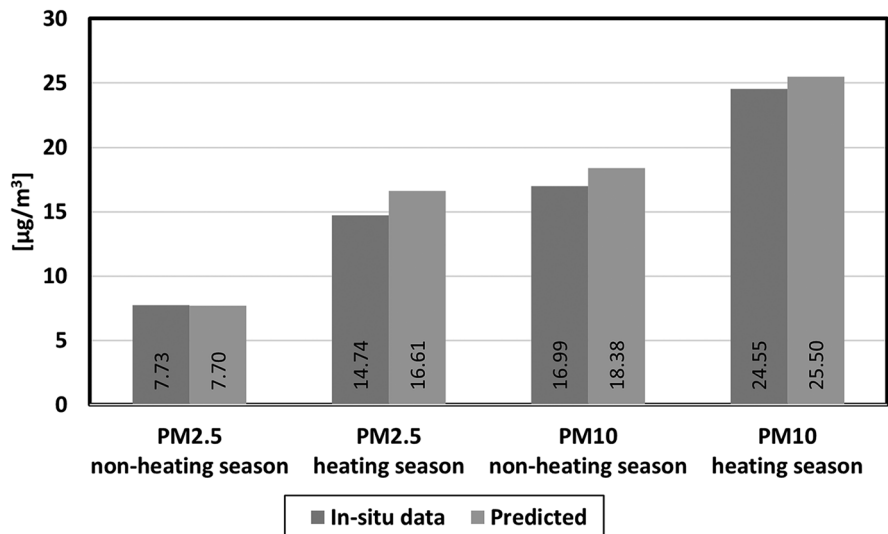
Fig. 12 All developed PM$_{2.5}$ and PM$_{10}$ models

Model	Pollutant	r	MAE	RMSE
Continental: non-heating season	PM$_{2.5}$	0.73	↓ 3.43	↓ 6.77
Continental: non-heating season	PM$_{10}$	0.69	⇨ 8.27	↑ 11.78
Mediterranean: non-heating season	PM$_{2.5}$	0.73	↓ 2.86	↓ 3.72
Mediterranean: non-heating season	PM$_{10}$	0.67	⇨ 7.44	↑ 14.11
Continental: heating season	PM$_{2.5}$	0.74	⇨ 9.28	↑ 12.36
Continental: heating season	PM$_{10}$	0.69	⇨ 10.43	↑ 14.51
Mediterranean: heating season	PM$_{2.5}$	0.82	↓ 3.73	↓ 5.41
Mediterranean: heating season	PM$_{10}$	0.73	⇨ 8.40	↑ 12.98
Croatia: non-heating season	PM$_{2.5}$	0.71	↓ 3.23	↓ 5.27
Croatia: non-heating season	PM$_{10}$	0.59	⇨ 7.64	↑ 13.33
Croatia: heating season	PM$_{2.5}$	0.73	↓ 6.30	⇨ 8.82
Croatia: heating season	PM$_{10}$	0.68	⇨ 10.23	↑ 15.02
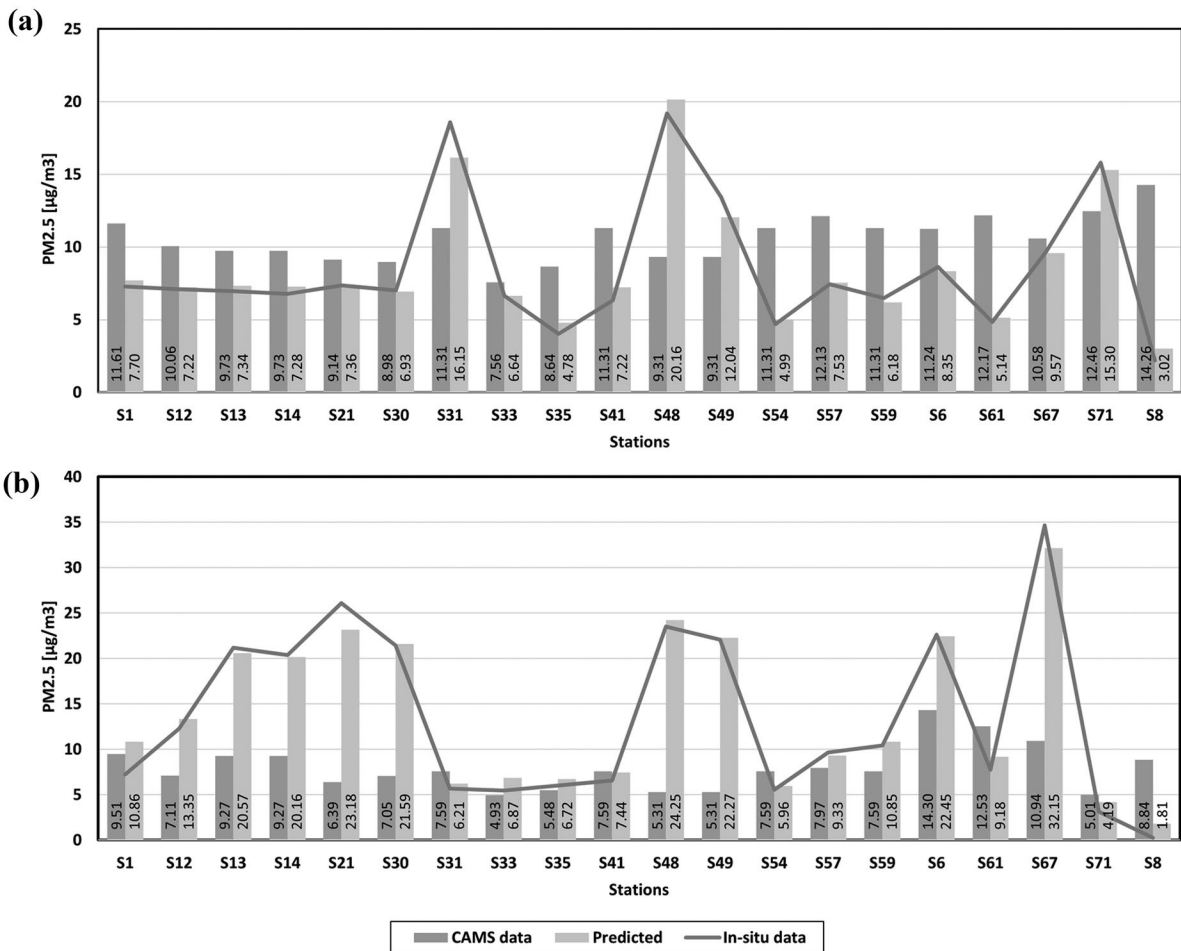
Fig. 13 PM$_{2.5}$ comparison between CAMS, predicted and in situ data for Croatia: **a** non-heating season and **b** heating season

Conclusions

This study followed a new approach to estimate ambient concentrations of $PM_{2.5}$ and PM_{10} from TRO-POMI and other open-source remote sensing data available on GEE. Therefore, on a non-heating and heating seasons time scale, the random forest machine learning method was successfully used to create moderate to high precision $PM_{2.5}$ and PM_{10} models for the Republic of Croatia. The spatial distribution of $PM_{2.5}$ concentrations during the heating season revealed significant variations in pollution between biogeographical regions, which motivated us to develop regional models as well.

Due to the insufficient number of ground monitoring stations in the Alpine region of Croatia, it was decided not to develop models for the Alpine region. However, the Continental region has sufficient ground monitoring stations and is the most polluted region in the country. As a result of their high correlation and little seasonal variation, all models trained for the Continental region are quite stable. The accuracy of $PM_{2.5}$ models ($0.73 < r < 0.74$) is higher than that of PM_{10} ($r = 0.69$). All developed models have a high correlation when it comes to the Mediterranean region, which is mainly influenced by the sea and strong winds. The accuracy of the models increases with the heating season, although pollution is lower in the heating season. The $PM_{2.5}$ model shows $r = 0.73$ for the non-heating season and $r = 0.82$ for the heating season. On the other hand, the PM_{10} model has $r = 0.67$ for non-heating seasons and $r = 0.73$ for heating seasons. Similarly to models for the Continental region, the accuracy of $PM_{2.5}$ models is higher than that of PM_{10}.

Developed models for predicting seasonal variations of $PM_{2.5}$ and PM_{10} on the whole territory of the Republic of Croatia show moderate to high accuracy. In particular, the $PM_{2.5}$ model for the non-heating season has $r = 0.71$ and $r = 0.73$ for the heating season. On the other hand, the PM_{10} model has $r = 0.59$ for the non-heating, and $r = 0.68$ for the heating season.

Comparison between in situ, predicted and CAMS $PM_{2.5}$ data have shown how CAMS is consistent, but unable to monitor sudden changes in the observed pollutant. Therefore, the developed models proved to be a better solution for monitoring atmospheric concentrations of $PM_{2.5}$ and PM_{10} over Croatia.

Most of the models developed slightly underestimate the actual values, making the predicted values appear slightly lower. However, even in places with low levels of pollution, all models have demonstrated a general ability to estimate $PM_{2.5}$ and PM_{10} levels. Additionally, all models can accurately identify all $PM_{2.5}$ and PM_{10} hotspots. The approach proposed by this study has great potential to be extended to a larger scale. However, manual cleaning of the data set can be challenging, especially in emergency situations of atmospheric pollution; thus, for future studies, we recommend automatization of the process which seems to be a key element of modeling. Also, future studies should focus on applying the regional models developed by this research in other Continental and Mediterranean regions in Europe.

Data availability The datasets generated and analyzed in this study are available from the corresponding author upon reasonable request.

Declarations

Ethical approval All authors have read, understood, and have complied as applicable with the statement on "Ethical responsibilities of Authors" as found in the Instructions for Authors.

Competing interests The authors declare no competing interests.

References

Ahmed, M., Xiao, Z., & Shen, Y. (2022). Estimation of ground PM2. 5 concentrations in Pakistan using convolutional neural network and multi-pollutant satellite images. *Remote Sensing, 14(7)*, 1735.

Ayturan, Y. A., Ayturan, Z. C., & Altun, H. O. (2018). Air pollution modelling with deep learning: A review. *International Journal of Environmental Pollution and Environmental Modelling, 1*(3), 58–62.

Biancofiore, F., Busilacchio, M., Verdecchia, M., Tomassetti, B., Aruffo, E., Bianco, S., & Di Carlo, P. (2017). Recursive neural network model for analysis and forecast of PM10 and PM2. 5. *Atmospheric Pollution Research, 8(4)*, 652–659.

Biau, G., & Scornet, E. (2016). A random forest guided tour. *TEST, 25*(2), 197–227.

Bodor, K., Szép, R., & Bodor, Z. (2022). The human health risk assessment of particulate air pollution (PM2. 5 and PM10) in Romania. *Toxicology Reports, 9*, 556–562.

Breiman, L. (2001). *Random Forests. Machine Learning, 45*(1), 5–32.

Chen, G., Li, S., Knibbs, L. D., Hamm, N. A., Cao, W., Li, T., & Guo, Y. (2018a). A machine learning method to estimate PM2. 5 concentrations across China with remote sensing, meteorological and land use information. *Science of the Total Environment, 636*, 52–60.

Chen, G., Wang, Y., Li, S., Cao, W., Ren, H., Knibbs, L. D., & Guo, Y. (2018b). Spatiotemporal patterns of PM10 concentrations over China during 2005–2016: a satellite-based estimation using the random forests approach. *Environmental pollution, 242*, 605–613.

Cichowicz, R., & Dobrzański, M. (2022). Analysis of air pollution around a CHP Plant: real measurements vs. computer simulations. *Energies 2022, 15*, 553. https://doi.org/10.3390/en15020553

Cichowicz, R., Wielgosiński, G., & Fetter, W. (2017). Dispersion of atmospheric air pollution in summer and winter season. *Environmental Monitoring and Assessment, 189*(12), 1–10.

De Fazio, R., Dinoi, L. M., De Vittorio, M., & Visconti, P. (2022). A sensor-based drone for pollutants detection in eco-friendly cities: Hardware design and data analysis application. *Electronics, 11*, 52. https://doi.org/10.3390/electronics11010052

EEA (2016). Biogeographical regions. Retrieved date September 11, 2022, from https://www.eea.europa.eu/data-and-maps/data/biogeographical-regions-europe-3

Faraji Ghasemi, F., Dobaradaran, S., Saeedi, R., Nabipour, I., Nazmara, S., Ranjbar Vakil Abadi, D., & Keshtkar, M. (2020). Levels and ecological and health risk assessment of PM2. 5-bound heavy metals in the northern part of the Persian Gulf. *Environmental Science and Pollution Research, 27(5)*, 5305–5313.

Fenger, J. (1999). Urban air quality. *Atmospheric Environment, 33*(29), 4877–4900.

Ghorani-Azam, A., Riahi-Zanjani, B., & Balali-Mood, M. (2016). Effects of air pollution on human health and practical measures for prevention in Iran. *Journal of research in medical sciences: the official journal of Isfahan University of Medical Sciences, 21*.

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., & Moore, R. (2017). Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sensing of Environment, 202*, 18–27.

Gulia, S., Nagendra, S. S., Khare, M., & Khanna, I. (2015). Urban air quality management-a review. *Atmospheric Pollution Research, 6*(2), 286–304.

Han, S., Kundhikanjana, W., Towashiraporn, P., & Stratoulias, D. (2022). Interpolation-based fusion of Sentinel-5P, SRTM, and regulatory-grade ground stations data for producing spatially continuous maps of PM2.5 concentrations nationwide over Thailand. *Atmosphere, 13(2)*, 161.

Hengl, T. (2018). Soil pH in H2O at 6 standard depths (0, 10, 30, 60, 100 and 200 cm) at 250 m resolution (Version v02) .

Hrvatska enciklopedija. (2021). Hrvatska. Retrieved date September 11, 2022, from https://www.enciklopedija.hr/natuknica.aspx?id=26390

Hu, K., Rahman, A., Bhrugubanda, H., & Sivaraman, V. (2017). HazeEst: Machine learning based metropolitan air pollution estimation from fixed and mobile sensors. *IEEE Sensors Journal, 17*(11), 3517–3525.

Huang, K., Xiao, Q., Meng, X., Geng, G., Wang, Y., Lyapustin, A., & Liu, Y. (2018). Predicting monthly high-resolution PM2. 5 concentrations with random forest model in the North China Plain. *Environmental pollution, 242*, 675–683.

Index. (2021). Hrvatska je na dnu EU po kvaliteti zraka. Retrieved date September 11, 2022, from https://www.index.hr/vijesti/clanak/hrvatska-je-u-vrhu-eu-po-broju-smrti-od-zagadjenog-zraka-to-nas-kosta-milijarde/2255229.aspx

ISZZ. (2022). Kvaliteta zraka u Republici Hrvatskoj. Retrieved date September 11, 2022, from http://iszz.azo.hr/iskzl/index.html

Kaplan, G., & Avdan, Z. Y. (2020). Space-borne air pollution observation from sentinel-5p tropomi: Relationship between pollutants, geographical and demographic data. *International Journal of Engineering and Geosciences, 5*(3), 130–137.

Kaplan, G., Avdan, Z. Y., & Avdan, U. (2019). Spaceborne nitrogen dioxide observations from the Sentinel-5P TRO-POMI over Turkey. *Multidisciplinary Digital Publishing Institute Proceedings, 18*(1), 4.

Larkin, Geddes, J. A., Martin, R. V., Xiao, Q., Liu, Y., Marshall, J. D., Brauer, M., & Hystad, P. (2017). Global Land Use Regression Model for Nitrogen Dioxide Air Pollution. *Environmental Science & Technology, 51*(12), 6957–6964. https://doi.org/10.1021/acs.est.7b01148

Lave, L. B., & Seskin, E. P. (2013). Air pollution and human health. *RFF Press*.

Leão, M. L. P., Zhang, L., & da Silva Júnior, F. M. R. (2022). Effect of particulate matter (PM2. 5 and PM10) on health indicators: climate change scenarios in a Brazilian metropolis. *Environmental Geochemistry and Health*, 1–12.

Li, T., Shen, H., Yuan, Q., & Zhang, L. (2020). Geographically and temporally weighted neural networks for satellite-based mapping of ground-level PM2.5. *ISPRS Journal of Photogrammetry and Remote Sensing, 167*, 178–188.

Li, T., Yang, Q., Wang, Y., & Wu, J. (2022). Joint estimation of PM2. 5 and O3 over China using a knowledge-informed neural network. *Geoscience Frontiers, 101499*.

Li, Y., Yuan, S., Fan, S., Song, Y., Wang, Z., Yu, Z., & Liu, Y. (2021). Satellite remote sensing for estimating PM2. 5 and Its Components. *Current Pollution Reports, 7(1)*, 72–87.

Lin, C., Li, Y., Yuan, Z., Lau, A. K., Li, C., & Fung, J. C. (2015). Using satellite remote sensing data to estimate the high-resolution distribution of ground-level PM2.5. *Remote Sensing of Environment, 156*, 117–128.

Mamić, L. (2021). Colevel over the Republic of Croatia using Sentinel-5P. *GIS Odyssey Journal, 1*(1), 61–82.

Mamić, L., Kaplan, G., & Gašparović, M. (2022). Estimating monthly levels of PM2.5 and PM10 in the Republic of Croatia from Sentinel-5P and assimilated datasets. *GIS Odyssey Journal, 2(2)*, 59–77.

Masih, A. (2019). Machine learning algorithms in air quality modeling. *Global Journal of Environmental Science and Management, 5*(4), 1–20.

Ning, G., Wang, S., Ma, M., Ni, C., Shang, Z., Wang, J., & Li, J. (2018). Characteristics of air pollution in different zones of Sichuan Basin, China. *Science of the Total Environment, 612*, 975–984.

Polichetti, G., Cocco, S., Spinali, A., Trimarco, V., & Nunziata, A. (2009). Effects of particulate matter (PM10, PM2.5 and PM1) on the cardiovascular system. *Toxicology, 261(1–2)*, 1–8.

Rosenfeld, D., Lahav, R., Khain, A., & Pinsky, M. (2002). The role of sea spray in cleansing air pollution over ocean via cloud processes. *Science, 297*(5587), 1667–1670.

Rybarczyk, Y., & Zalakeviciute, R. (2018). Machine learning approaches for outdoor air quality modelling: A systematic review. *Applied Sciences, 8*(12), 2570.

Samad, A., Alvarez Florez, D., Chourdakis, I., & Vogt, U. (2022). Concept of using an unmanned aerial vehicle (UAV) for 3D investigation of air quality in the atmosphere—example of measurements near a roadside. *Atmosphere, 13*, 663. https://doi.org/10.3390/atmos13050663

Saurabh Sonwani, S. S., & Vandana Maurya, V. M. (2019). Impact of air pollution on the environment and economy. *Air pollution: Sources, impacts and controls* (pp. 113–134). CAB International.

Shahriyari, H. A., Nikmanesh, Y., Jalali, S., Tahery, N., Zhiani Fard, A., Hatamzadeh, N., & Mohammadi, M. J. (2022). Air pollution and human health risks: mechanisms and clinical manifestations of cardiovascular and respiratory diseases. *Toxin Reviews, 41(2)*, 606–617.

Shao, Y., Ma, Z., Wang, J., & Bi, J. (2020). Estimating daily ground-level PM2.5 in China with random-forest-based spatiotemporal kriging. *Science of The Total Environment, 740*, 139761.

Son, R., Kim, H. C., Yoon, J. H., & Stratoulias, D. (2022). Estimation of surface Pm2.5 concentrations from atmospheric gas species retrieved from tropomi using deep learning: impacts of fire on air pollution over Thailand. Available at SSRN 4255502.

Stafoggia, M., Bellander, T., Bucci, S., Davoli, M., De Hoogh, K., De'Donato, F., & Schwartz, J. (2019). Estimation of daily PM10 and PM2.5 concentrations in Italy, 2013–2015, using a spatiotemporal land-use random-forest model. *Environment international, 124*, 170–179.

Stevens, C. J., Bell, J. N. B., Brimblecombe, P., Clark, C. M., Dise, N. B., Fowler, D., & Wolseley, P. A. (2020). The impact of air pollution on terrestrial managed and natural vegetation. *Philosophical Transactions of the Royal Society A, 378(2183)*, 20190317.

Sun, Y., Zeng, Q., Geng, B., Lin, X., Sude, B., & Chen, L. (2019). Deep learning architecture for estimating hourly ground-level PM 2.5 using satellite remote sensing. *IEEE Geoscience and Remote Sensing Letters, 16(9)*, 1343–1347.

Van Roode, S., Ruiz-Aguilar, J. J., González-Enrique, J., & Turias, I. J. (2019). An artificial neural network ensemble approach to generate air pollution maps. *Environmental Monitoring and Assessment, 191*(12), 1–15.

Wang, Y., Yuan, Q., Li, T., Tan, S., & Zhang, L. (2021). Full-coverage spatiotemporal mapping of ambient PM2.5 and PM10 over China from Sentinel-5P and assimilated datasets: Considering the precursors and chemical compositions. *Science of The Total Environment, 793*, 148535.

Witten, I. H., Frank, E., Trigg, L. E., Hall, M. A., Holmes, G., & Cunningham, S. J. (1999). Weka: Practical machine learning tools and techniques with Java implementations.

Xiao, Q., Ma, Z., Li, S., & Liu, Y. (2015). The impact of winter heating on air pollution in China. *PLoS ONE, 10*(1), e0117311.

Yang, L., Xu, H., & Yu, S. (2020). Estimating PM2.5 concentrations in Yangtze River Delta region of China using random forest model and the Top-of-Atmosphere reflectance. *Journal of Environmental Management, 272*, 111061.

Yunesian, M., Rostami, R., Zarei, A., Fazlzadeh, M., & Janjani, H. (2019). Exposure to high levels of PM2.5 and PM10 in the metropolis of Tehran and the associated health risks during 2016–2017. *Microchemical Journal, 150*, 104174.

Zhang, Y., & Li, Z. (2015). Remote sensing of atmospheric fine particulate matter (PM2.5) mass concentration near the ground from satellite observation. *Remote Sensing of Environment, 160*, 252–262.

Zhao, C., Wang, Q., Ban, J., Liu, Z., Zhang, Y., Ma, R., & Li, T. (2020). Estimating the daily PM2.5 concentration in the Beijing-Tianjin-Hebei region using a random forest model with a 0.01°× 0.01° spatial resolution. *Environment international, 134*, 105297.

Zoran, M. A., Savastru, R. S., Savastru, D. M., & Tautan, M. N. (2020). Assessing the relationship between surface levels of PM2.5 and PM10 particulate matter impact on COVID-19 in Milan, Italy. *Science of the Total Environment, 738*.