GIS Odyssey Journal

2022 • Volume 2 • Number 2

https://doi.org/10.57599/gisoj.2022.2.2.59

Luka Mamić[1], Gordana Kaplan[2], Mateo Gašparović[3]

FUSING MULTIPLE OPEN-SOURCE REMOTE SENSING DATA TO ESTIMATE $PM_{2.5}$ AND PM_{10} MONTHLY CONCENTRATIONS IN CROATIA

Abstract: The objective of this study is to create a methodology 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 on a monthly basis for June, July and August which are considered as months of non-heating season in Croatia, and December, January and February, which, on the other hand, are considered as months of the heating season. Furthermore, machine learning algorithms were employed in this study to build models that can accurately identify air quality. The proposed method uses open-source remote sensing data accessible on the GEE platform, with in-situ data from Croatian National Network for Continuous Air Quality Monitoring as ground truth data. A common thing for all developed monthly models is that the predicted values slightly underestimate the actual ones and appear slightly lower. However, all models have shown the general ability to estimate $PM_{2.5}$ and PM_{10} levels, even in areas without high pollution. All developed models show moderate to high correlation between in-situ and estimated $PM_{2.5}$ and PM_{10} values, with overall better results for $PM_{2.5}$ than for PM_{10} concentrations. Regarding $PM_{2.5}$ models, the model with the highest correlation ($r = 0.78$) is for January. The PM_{10} model with the highest correlation ($r = 0.79$) is for December. All things considered, developed models can effectively detect all $PM_{2.5}$ and PM_{10} hotspots.

Keywords: air quality, TROPOMI, machine learning, $PM_{2.5}$, PM_{10}, remote sensing

Received: 2 November 2022; accepted: 14 December 2022

[1] Sapienza University of Rome, Department of Civil, Building and Environmental Engineering, Rome, Italy, ORCID ID: https://orcid.org/0000-0001-9877-0174, email: luka.mamic@uniroma1.it
[2] Eskisehir Technical University, Institute of Earth and Space Sciences, Eskisehir, Turkey, ORCID ID: https://orcid.org/0000-0001-7522-9924, email: kaplangorde@gmail.com
[3] University of Zagreb, Faculty of Geodesy, Chair of Photogrammetry and Remote Sensing, Zagreb, Croatia, ORCID ID: https://orcid.org/0000-0003-2345-7882, email: mateo.gasparovic@geof.unizg.hr

Introduction

The World Health Organization (WHO) estimates that 99 percent of the world's population breaths air that exceeds their guidelines. The cumulative impacts of air pollution result in over seven million premature deaths annually, making it a serious threat to both human health and the environment (Air pollution, URL 1).

Since it has been shown that particulate matter (PM) is dangerous to human health (Pope III & Dockery, 2006, Anderson et al., 2012; Kim et al., 2015; Kumar et al., 2022), it has received significant attention among air pollutants. Based on their aerodynamic diameter, airborne particles are classified as coarse (i.e., PM_{10}, particles with an aerodynamic diameter less than or equal to 10 μm) or fine (i.e., $PM_{2.5}$, particles with an aerodynamic diameter less than or equal to 2.5 μm).

Even a 10 $g/\mu m^3$ increase in PM_{10} has been shown to increase the chance of hospitalization for myocardial infarction. Furthermore, even a few hours of exposure to high levels of $PM_{2.5}$ increases the risk of myocardial infarction in a high-risk group (Polichetti et al., 2009).

Source and chemical composition of $PM_{2.5}$ and PM_{10} changes in time and space due to various parameters, such as human activity, natural hazards, temperature changes and others (Kelly & Fussell, 2012; Clemente et al., 2022; Gautam et al., 2022; Singh et al., 2022). Therefore, ground stations are commonly used to monitor $PM_{2.5}$ and PM_{10} concentrations but are limited to local areas and are unable to show a broad view (Li et al., 2020). That being said, predicting $PM_{2.5}$ and PM_{10} atmospheric concentrations over larger areas still remains a challenge. However, approaches based on remote sensing have lately been developed. Several studies have been done so far on estimating ambient PM concentrations ($PM_{2.5}$ and PM_{10}) using observations from remote sensing satellites (Li et al., 2021; Wang et al., 2021; Han et al., 2022).

Li et al. (2021) provided a review of the two-step methods for estimating $PM_{2.5}$, which first retrieve the aerosol optical depth (AOD) and then estimate $PM_{2.5}$ from the AOD with other supplemental data containing the temporal or spatial variation impact on $PM_{2.5}$. The new approach for estimating PMs comes with the launch of the first Copernicus mission on monitoring air quality, Sentinel-5P TROPOMI, which offers daily measured ozone (O_3), methane (CH_4), formaldehyde (HCHO), carbon monoxide (CO), nitrogen oxide (NO_2), sulphur dioxide (SO_2), and aerosol index (AI) in the atmosphere. Therefore, Wang et al. (2021) proposed a methodology for $PM_{2.5}$ and PM_{10} estimation over China using TROPOMI and assimilated datasets, and compared it with the traditional approaches which are using AOD data. Their conclusion was that the TROPOMI data have great potential in this topic and fix the issue due to the missing AOD data. More recently, Han et al. (2022) created spatially continuous maps of $PM_{2.5}$ concentrations over Thailand using TROPOMI, elevation, and regulatory grade ground station data.

This research follows a similar approach fusing TROPOMI with multiple open-source remote sensing data from the Google Earth Engine (GEE) platform to create a new method for accurately estimating atmospheric concentrations of $PM_{2.5}$ and PM_{10}

over Croatia for six chosen months. The new approach proposed by this study combines air pollution, meteorological and geographical data to create new parameters in order to answer the questions of $PM_{2.5}$ and PM_{10} composition in Croatia and improve the stability and accuracy of developed models. Due to the variation of air pollutants between the non-heating and heating seasons (Xiao et al., 2015; Cichowicz et al., 2017) modelling was done for June, July, and August which are considered as months of the non-heating seasons in Croatia, and December, January, and February, which, on the other hand, are considered as months of the heating season. Therefore, the objectives of this study are to develop $PM_{2.5}$ and PM_{10} models for months of non-heating and heating season in Croatia.

Furthermore, Random Forest machine learning algorithm, widely used when estimating $PM_{2.5}$ and PM_{10} (Stafoggia et al., 2019; Shao et al., 2020; Zhao et al., 2020; Yang et al., 2020), was employed in this study to build models that can accurately identify air quality. The results showed that new approach developed by this study estimates $PM_{2.5}$ and PM_{10} accurately and, most importantly, can be easily adopted to new study areas.

Materials and methods

In-situ data. There are 71 ground stations in the Croatian National Network for Continuous Air Quality Monitoring (Fig. 1.), but only 31 measure $PM_{2.5}$ and 54 measure PM_{10} in the atmosphere.

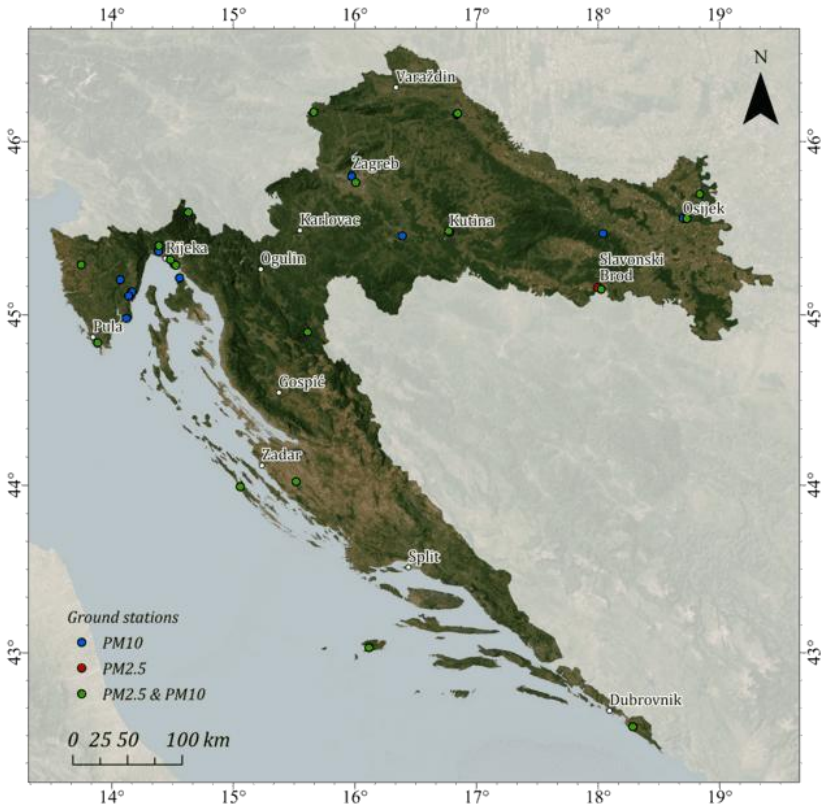


Fig. 1. Croatian National Network for Continuous Air Quality Monitoring
Source: own study

Every hour, automatic stations measure $PM_{2.5}$ and PM_{10} as $\mu g/m^3$. All data is stored and publicly accessible on the Croatian National Network for Continuous Air Quality Monitoring's official website (Kvaliteta zraka u Republici Hrvatskoj, URL 2).

$PM_{2.5}$ and PM_{10} raw hourly data in Excel format (.xls) were downloaded for all stations for June, July, August, and December 2021, as well as for January and February 2022. Furthermore, some data for the observed time range were missing or invalid, so finally data from 20 ground stations were used to estimate $PM_{2.5}$ for all months and PM_{10} from 30 for June, July and August and from 32 for December, January and February. Additionally, ground measurements that were too high or too low compared to the rest of the dataset were removed. Finally, in-situ data was matched to remote sensing data.

Sentinel-5P TROPOMI data. The main data used to estimate the $PM_{2.5}$ and PM_{10} values in this study was TROPOMI data from GEE.

Every day, the Sentinel-5P mission collects data that are used to monitor and forecast air quality, climatic conditions, ozone, and ultraviolet radiation levels. It is the satellite mission launched on October 13, 2017, carrying the Tropospheric Monitoring Instrument (TROPOMI), and it is the first in a series of atmospheric observing systems within Copernicus, the European Union's Earth observation program (Kleipool et al., 2018).

TROPOMI can observe and monitor air pollution from hotspots such as large cities, and industrial areas (Borsdorff et al., 2018; Kaplan & Avdan, 2020). Furthermore, the main topic of this study is to explore the potential of Sentinel-5P data to estimate $PM_{2.5}$ and PM_{10}, both of which are important in determining air quality.

Preprocessed L3 TROPOMI data of AI, CO, HCHO, NO_2, O_3, SO_2 were used for modelling. CH_4 data is missing for all observed periods, and SO_2 data is missing for the winter months (December 2021, January, and February 2022). Furthermore, due to the increased number of cloudy days during the winter months, some CO, HCHO, and NO_2 data are missing. Moreover, due to the reported systematic error for measuring the AI until July 2021, AI data from June was also excluded from the modelling. The unit of TROPOMI measured CO, HCHO, NO_2, O_3, and SO_2 is mol/m^2.

Based on the fact that TROPOMI captures images over Croatia every day between 10:00 AM and 1:00 PM CET, the TROPOMI data were matched with the in-situ data for the precise hour.

Assimilated data. Other meteorological and geographical data from the GEE platform were used to improve the models.

Through its Global Forecast System, the National Oceanic and Atmospheric Administration (NOAA) provides a dataset consisting of selected model outputs as gridded forecast variables (GFS). The 384-hour forecast includes a 3-hour forecast interval and a 6-hour temporal resolution (i.e., updated four times daily). It is a coupled model composed of an atmospheric model, an ocean model, a land/soil model, and a sea ice model, all of which work together to create a realistic representation of meteorological conditions (GFS: Global Forecast System 384-Hour Predicted Atmosphere Data, URL 3).

GFS data used in this study 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).

Elevation (DEM) and slope for the ground station locations were extracted from NASA's Shuttle Radar Topography Mission (SRTM) with a spatial resolution of 30 meters (Farr et al., 2007).

Soil pH data at ground level for all stations were extracted from the map made by Hengl in 2018 with a resolution of 250 m.

Parameters. The original TROPOMI parameters and the assimilated datasets are shown in Table 1.

Table 1. Original parameters

Parameter	Description	Source	Unit	Temporal resolution	Spatial resolution
AI	Aerosol index	TROPOMI	/	daily	1113.2 m
CO	Carbon monoxide	TROPOMI	mol/m^2	daily	1113.2 m
HCHO	Formaldehyde	TROPOMI	mol/m^2	daily	1113.2 m
NO_2	Nitrogen dioxide	TROPOMI	mol/m^2	daily	1113.2 m
O_3	Ozone	TROPOMI	mol/m^2	daily	1113.2 m
SO_2	Sulphur dioxide	TROPOMI	mol/m^2	daily	1113.2 m
LST	Land surface temperature	NOAA	°C	6 hours	27 890 m
HUM	Specific humidity	NOAA	kg/kg	6 hours	27 890 m
U-WIND	Eastward wind	NOAA	m/s	6 hours	27 890 m
V-WIND	Northward wind	NOAA	m/s	6 hours	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

Source: own study

Using the main data (TROPOMI) and the assimilated data, new parameters for the modelling were created to improve the stability of the models (Table 2).

The parameters listed above were created based on the things they have in common (i.e., NO_2 and SO_2 have the same chemical structure, are both inorganic and react similarly). Furthermore, some were created solely as normalized ratios, while others were merged based on their source (i.e., WHT is a parameter composed of wind components, humidity, and temperature).

Table 2. Created parameters

Parameter
$(NO_2+SO_2)/(NO_2-SO_2)$
NO_2/SO_2
HCHO/CO
(CO+HCHO)/CO-HCHO)
$O_3/(NO_2+SO_2+CO)$
$AI*(NO_2/SO_2)$
$O_3/(NO_2/SO_2)$
$O_3/((NO_2+SO_2)/(NO_2-SO_2))$
$SQRT(1/(NO_2+SO_2+O_3))$
(AI+HUM)/(AI-HUM)
(AI+DEM)/(AI-DEM)
$(CO+NO_2)/CO-NO_2)$
$(CO+O_3)/(CO-O_3)$
WIND[1]
WHT[2]
(WHT+AI)/(WHT-AI)
$(WHT+O_3)/(WHT-O_3)$
$(CO+SO_2)/(CO-SO_2)$
S5[3]
(S5+WHT)/(S5-WHT)
(S5+WIND)/(S5-WIND)

[1](U-WIND + V-WIND)/2, [2](((U-WIND + V-WIND)/2)+HUM+LST)/3,
[3](AI+CO+HCHO+NO_2+O_3+SO_2)/6

Source: own study

Modelling with Weka. The Waikato Environment for Knowledge Analysis (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).

Weka is open-source software with tools for data preprocessing, classification, regression, clustering, association rules, and visualization that is licensed under the GPL (General Public License).

Before modelling, selecting only the best parameters for each model was necessary. This was done using Weka's Classifier subset evaluator tool for the Random Forest algorithm with a percentage split of 70. Once the best parameters were found, all remaining were removed, and the Random Forest classifier was used with data split into training and testing portions of 70% and 30%, respectively.

The Random Forest classifier comprises a group of tree classifiers, each of which is generated using a random vector sampled independently from the input vector. Each tree casts a unit vote for the most popular class to classify an input vector (Pal, 2005).

The basic idea behind Random Forest is a simple but powerful one – the wisdom of crowds. The Random Forest model works so well because a large number of relatively

uncorrelated models (trees) acting as a committee will outperform any of the individual constituent models (Understanding Random Forest, URL 4).

The model was saved once it was developed. All models in this study are created using the same procedure as described above.

The number of instances and parameters used to create each model is shown in Table 3 below.

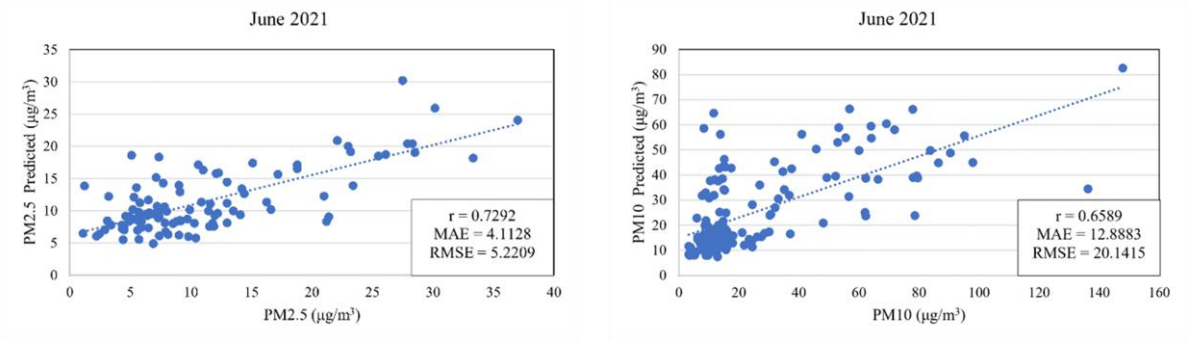
Table 3. Number of instances used to create each model

Model	Pollutant	Number of instances	Number of parameters
June	$PM_{2.5}$	350	12
June	PM_{10}	484	5
July	$PM_{2.5}$	271	4
July	PM_{10}	364	18
August	$PM_{2.5}$	331	11
August	PM_{10}	459	10
December	$PM_{2.5}$	178	6
December	PM_{10}	268	6
January	$PM_{2.5}$	280	7
January	PM_{10}	478	6
February	$PM_{2.5}$	232	10
February	PM_{10}	363	7

Source: own study

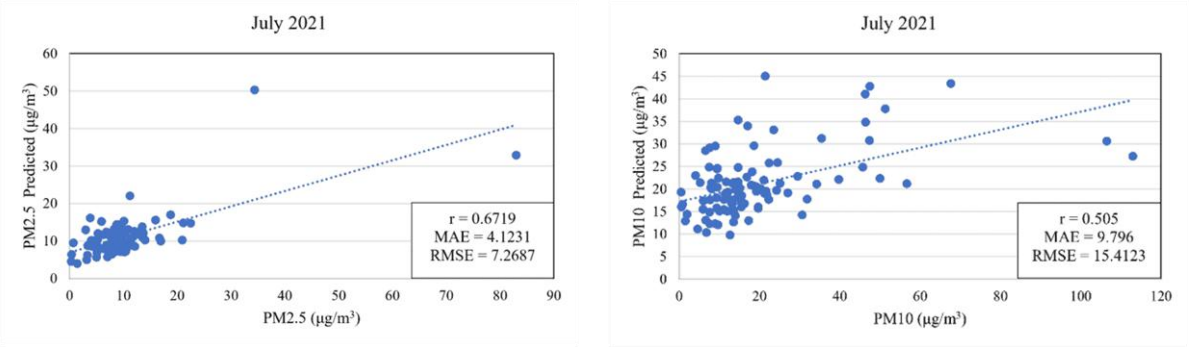
Results and discussion

The monthly $PM_{2.5}$ and PM_{10} models were developed, as noted before, using the Random Forest algorithm with a 70% split, and are shown in Fig. 2.–7. with their correlation coefficient (r), mean absolute error (MAE), and root mean squared error (RMSE).



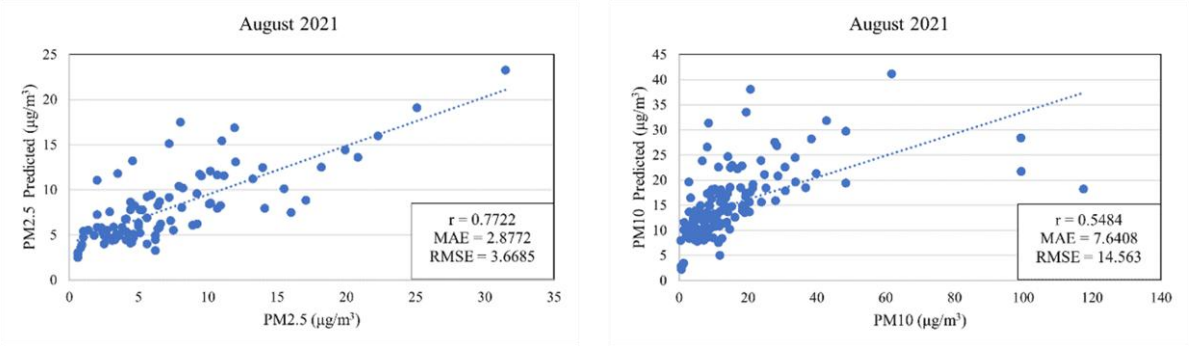
(a) (b)

Fig. 2. June prediction models for: (a) $PM_{2.5}$ (b) PM_{10}
Source: own study



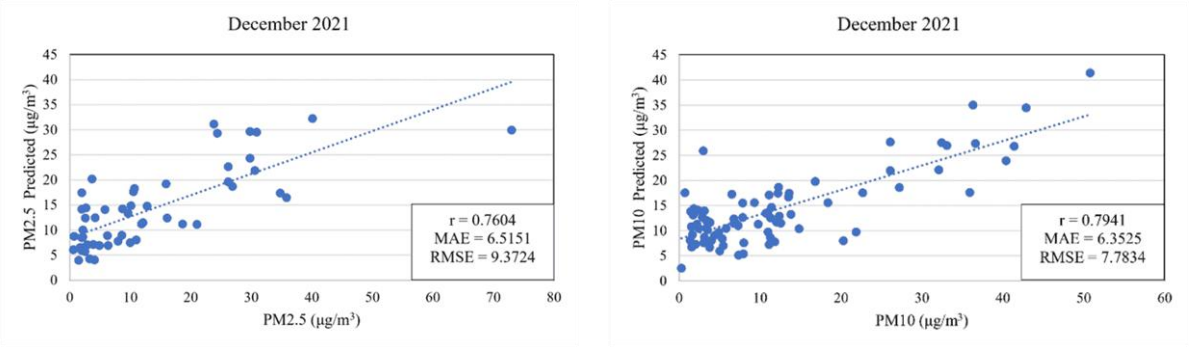
(a) (b)

Fig. 3. July prediction models for: (a) $PM_{2.5}$ (b) PM_{10}
Source: own study



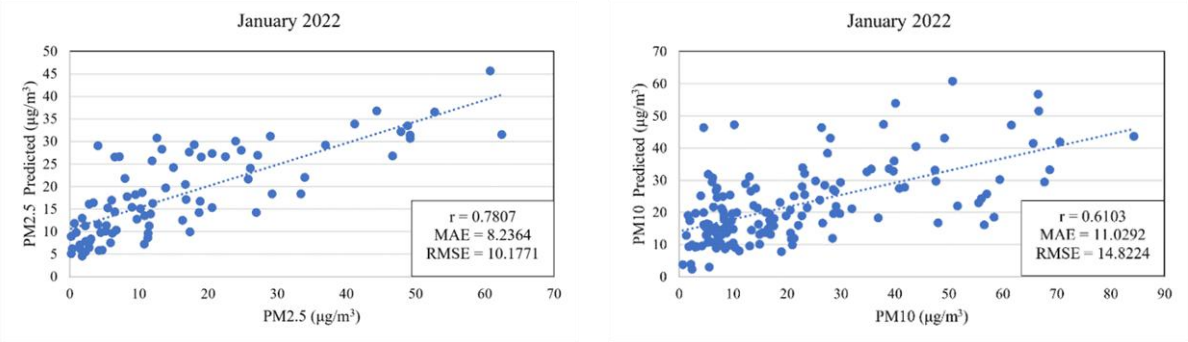
(a) (b)

Fig. 4. August prediction models for: (a) $PM_{2.5}$ (b) PM_{10}
Source: own study



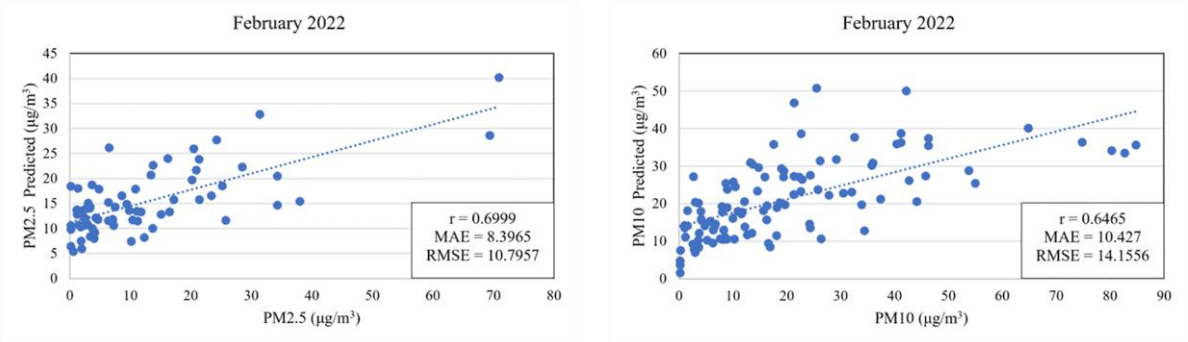
(a) (b)

Fig. 5. December prediction models for: (a) $PM_{2.5}$ (b) PM_{10}
Source: own study



(a) (b)

Fig. 6. January prediction models for: (a) $PM_{2.5}$ (b) PM_{10}
Source: own study



(a) (b)

Fig. 7. February prediction models for: (a) $PM_{2.5}$ (b) PM_{10}
Source: own study

All monthly models give moderate to high correlation, with overall better results for $PM_{2.5}$ than for PM_{10}. Regarding $PM_{2.5}$ models, the one with the lowest correlation is for July with r = 0.67 (MAE = 4.12 µg/m³, RMSE = 7.27 µg/m³), and on the contrary, the one with the highest correlation is for January with r = 0.78 (MAE = 8.24 µg/m³, RMSE = 10.18 µg/m³). On the other hand, the PM_{10} model with the lowest correlation is also the one for July with r = 0.51 (MAE = 9.80 µg/m³, RMSE = 15.41 µg/m³), and the highest for December with r = 0.79 (MAE = 6.35 µg/m³, RMSE = 7.78 µg/m³).

The attributes chosen to build $PM_{2.5}$ and PM_{10} monthly models can give us a better understanding of each exact model. The selected parameters used to develop mentioned models and estimate $PM_{2.5}$ and PM_{10} monthly values are shown in Table 4.

What can be noticed is that the PM_{10} model for July used the highest number of parameters to develop, 18 to be exact, and it came out as the one with the lowest overall correlation. Furthermore, the $PM_{2.5}$ model for July used only four parameters, and even its accuracy is the lowest among $PM_{2.5}$ monthly models, it is still better than most of those for PM_{10}. The most common number of used parameters ranges from five to 12.

Table 4. Parameters used to develop $PM_{2.5}$ and PM_{10} monthly models

Pollutant	Month	Parameters
$PM_{2.5}$	June	O_3, SO_2, HUM, U-WIND, DEM, (CO+HCHO)/(CO-HCHO), O_3/(NO_2+SO_2+CO), O_3/(NO_2/SO_2), SQRT(1/(NO_2+SO_2+O_3)), (CO+O_3)/(CO-O_3), WIND[1], (CO+SO_2)/(CO-SO_2).
PM_{10}	June	HUM, SOIL_pH, DEM, SLOPE, SQRT(1/(NO_2+SO_2+O_3)).
$PM_{2.5}$	July	SO_2, DEM, NO_2/SO_2, O_3/((NO_2+SO_2)/(NO_2-SO_2)).
PM_{10}	July	HUM, U-WIND, SOIL_pH, DEM, SLOPE, (NO_2+SO_2)/(NO_2-SO_2), NO_2/SO_2, AI*(NO_2/SO_2), O_3/(NO_2/SO_2), O_3/((NO_2+SO_2)/(NO_2-SO_2)), (AI+HUM)/(AI-HUM), (AI+DEM)/(AI-DEM), (CO+NO_2)/(CO-NO_2), (CO+O_3)/(CO-O_3), WIND, (WHT[2]+AI)/(WHT-AI), (WHT+O_3)/(WHT-O_3), (S5[3]+WIND)/(S5-WIND).
$PM_{2.5}$	August	O_3, SO_2, LST, DEM, SLOPE, (CO+HCHO)/(CO-HCHO), O_3/(NO_2+SO_2+CO), O_3/((NO_2+SO_2)/(NO_2-SO_2)), (CO+NO_2)/(CO-NO_2), (CO+O_3)/(CO-O_3), WHT, (WHT+AI)/(WHT-AI).
PM_{10}	August	V-WIND, SOIL_pH, DEM, SLOPE, O_3/(NO_2+SO_2+CO), SQRT(1/(NO_2+SO_2+O_3), (CO+NO_2)/CO-NO_2), (WHT+AI)/(WHT-AI), (CO+SO_2)/(CO-SO_2), (S5+WIND)/(S5-WIND).
$PM_{2.5}$	December	O_3, U-WIND, DEM, SLOPE, HCHO/CO, (CO+HCHO)/(CO-HCHO).
PM_{10}	December	NO_2, O_3, DEM, SLOPE, (CO+O_3)/(CO-O_3), (WHT+AI)/(WHT-AI).
$PM_{2.5}$	January	CO, NO_2, HUM, SLOPE, HCHO/CO, (CO+NO_2)/(CO-NO_2), (CO+O_3)/(CO-O_3).
PM_{10}	January	SOIL_pH, DEM, SLOPE, WIND, WHT, (WHT+O_3)/(WHT-O_3).
$PM_{2.5}$	February	CO, NO_2, O_3, LST, HUM, V-WIND, SOIL_pH, (AI+DEM)/(AI-DEM), WHT, (WHT+O_3)/(WHT-O_3).
PM_{10}	February	AI, O_3, HUM, SOIL_pH, DEM, SLOPE, (WHT+AI)/(WHT-AI).
$PM_{2.5}$	June	O_3, SO_2, HUM, U-WIND, DEM, (CO+HCHO)/(CO-HCHO), O_3/(NO_2+SO_2+CO), O_3/(NO_2/SO_2), SQRT(1/(NO_2+SO_2+O_3)), (CO+O_3)/(CO-O_3), WIND[1], (CO+SO_2)/(CO-SO_2).

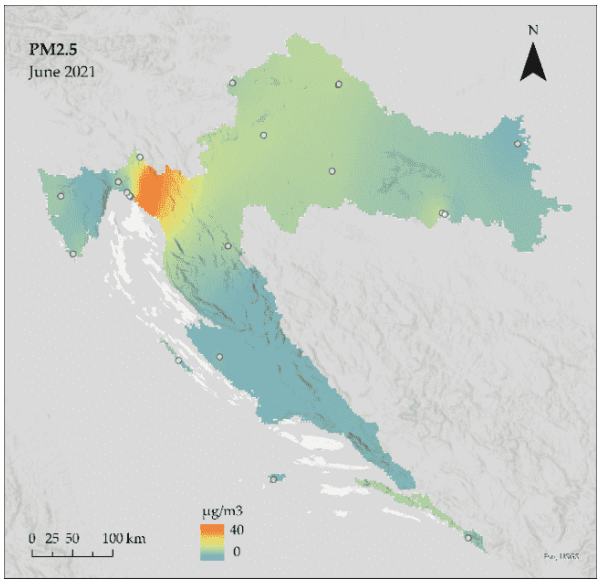
[1](U-WIND + V-WIND)/2, [2](((U-WIND + V-WIND)/2)+HUM+LST)/3, [3](AI+CO+HCHO+NO_2+O_3+SO_2)/6

Source: own study

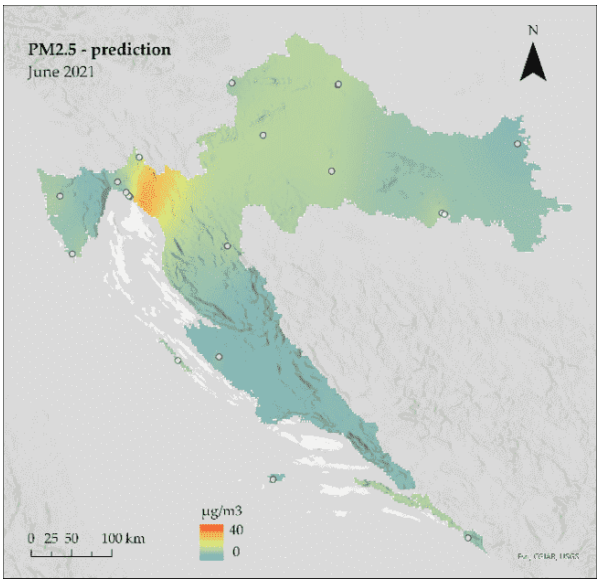
Compared with the study done for China by Wang et al. (2021), our study provides a more straightforward approach and uses fewer parameters, still providing satisfactory results. Wang et al. (2021) used 30 parameters from various sources (TROPOMI, GEOS-FP, MODIS, Open Street Map, and GPW) and showed that SO_2, NO_2, and wind components were the most important parameters. On the other hand, Han et al. (2022) used only TROPOMI and NASA SRTM elevation data in their study for Thailand. Unlike Wang et al. (2021), Han et al. (2022), have shown NO_2 and SO_2 as the most insignificant variables.

To get a visual insight between the in-situ and predicted data, the $PM_{2.5}$ and PM_{10} monthly interpolation maps were made using minimum curvature spline technique

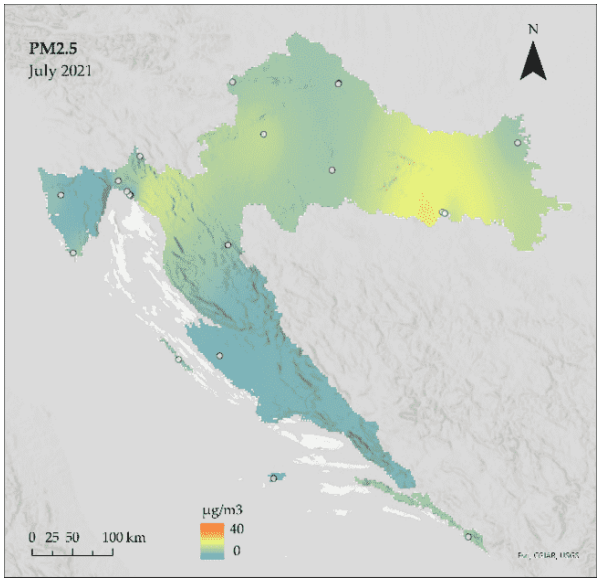
(Fig. 8.–11.). The ground stations used for modelling are indicated as point data on the map.



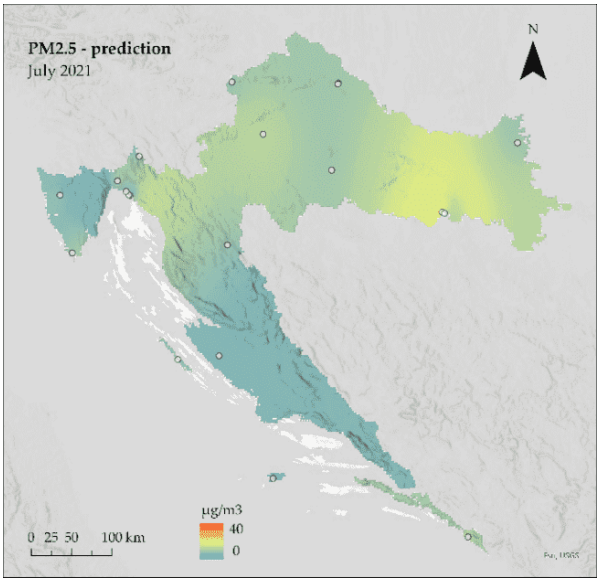
(a)



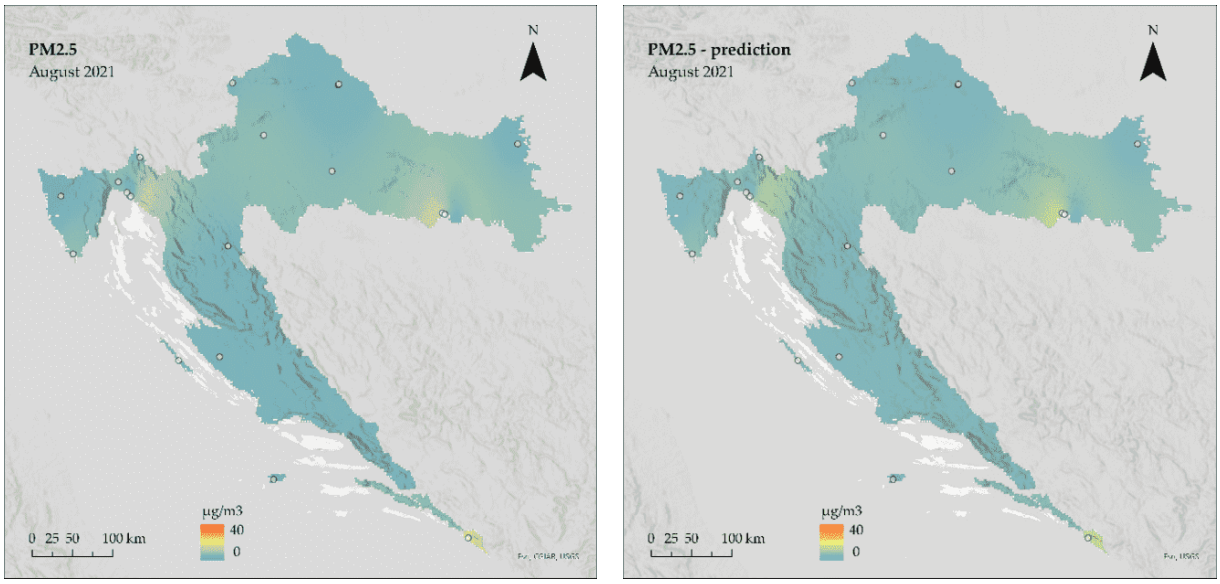
(b)



(c)



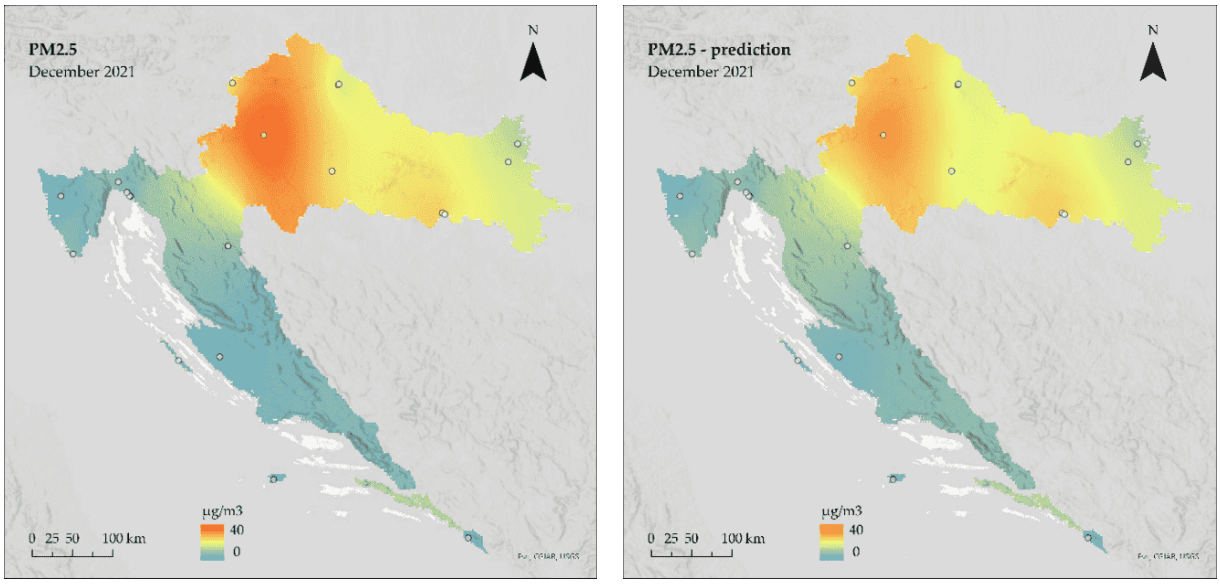
(d)



(e) (f)

Fig. 8. Interpolated monthly $PM_{2.5}$ concentrations: (a) June from in-situ data (b) June from remote sensing data (c) July from in-situ data (d) July from remote sensing data (e) August from in-situ data (f) August from remote sensing data.
Source: own study

We can easily spot two hotspots for the months of the non-heating season for $PM_{2.5}$, especially in June and July. The first one is located by the sea in north-western part of the country, around the urban area of the city Rijeka, and the second one in south-east of continental part of the country close to the border. There are only noticeable differences between the ground and the predicted data, where the predicted values appear to be slightly lower than the actual values (most visible for June).



(a) (b)

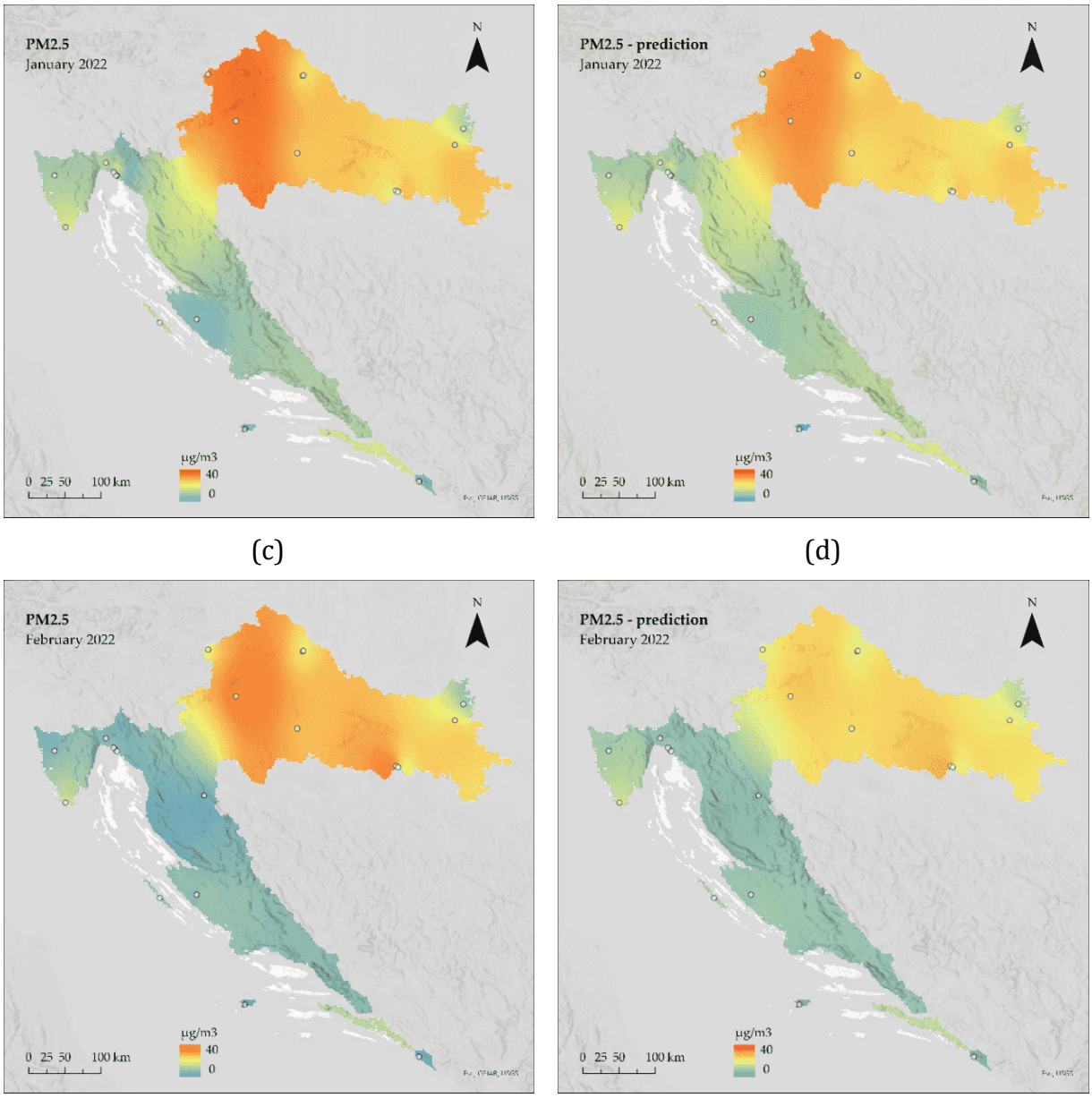
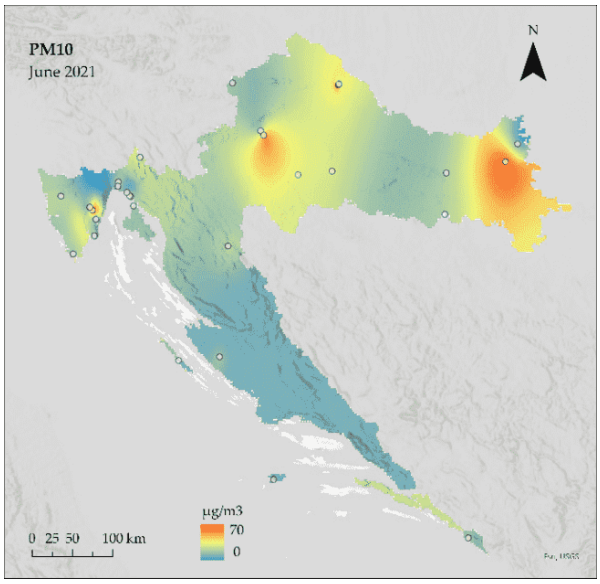
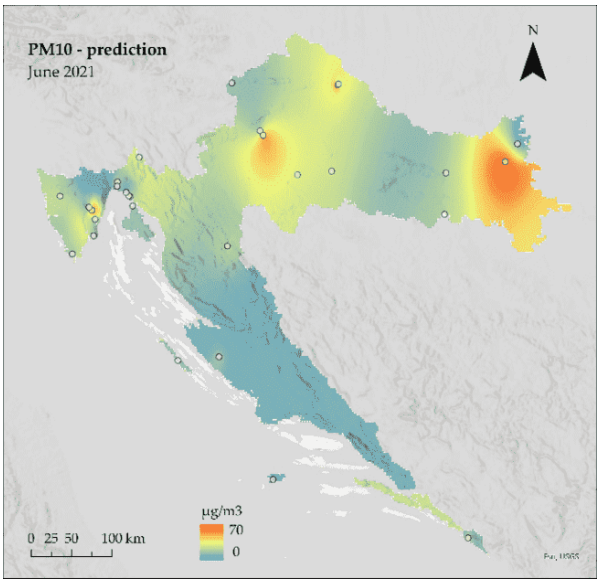


Fig. 9. Interpolated monthly $PM_{2.5}$ concentrations: (a) December from in-situ data (b) December from remote sensing data (c) January from in-situ data (d) January from remote sensing data (e) February from in-situ data (f) February from remote sensing data Source: own study

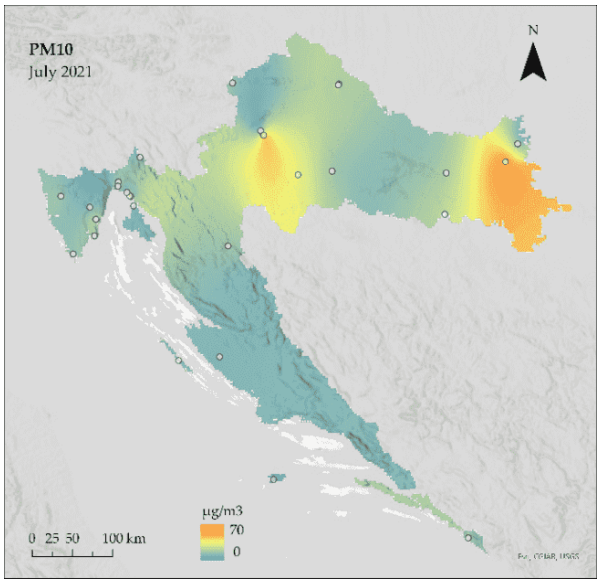
For the months of heating season, it is clearly visible that all the highest $PM_{2.5}$ values are in the east of the country (continental part of Croatia). Wang et al. (2021) observed similar seasonal changes and linked them to heating emissions and diverse meteorological conditions. The predicted values slightly underestimated the actual ones and appeared slightly lower.


PM10
June 2021
N
µg/m3
70
0
0 25 50 100 km

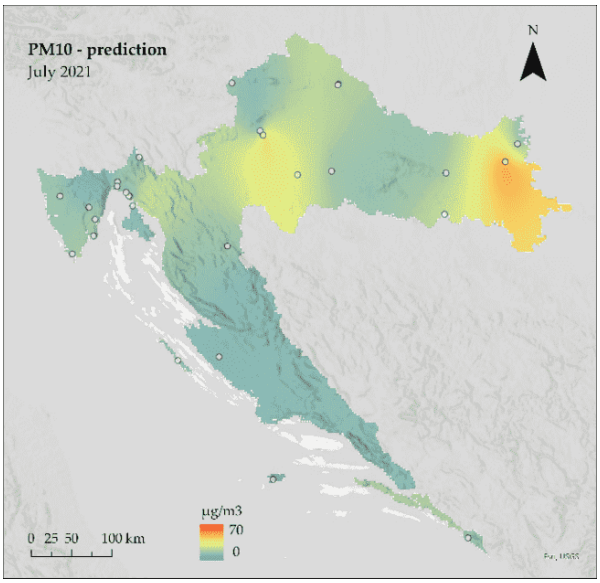
(a)


PM10 - prediction
June 2021
N
µg/m3
70
0
0 25 50 100 km

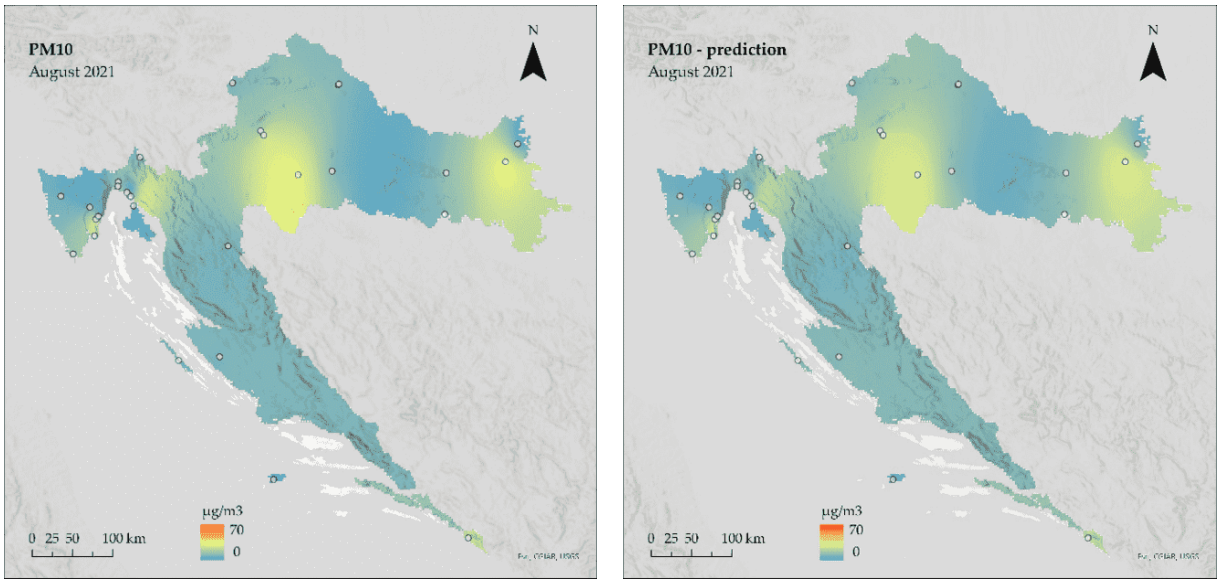
(b)


PM10
July 2021
N
µg/m3
70
0
0 25 50 100 km

(c)


PM10 - prediction
July 2021
N
µg/m3
70
0
0 25 50 100 km

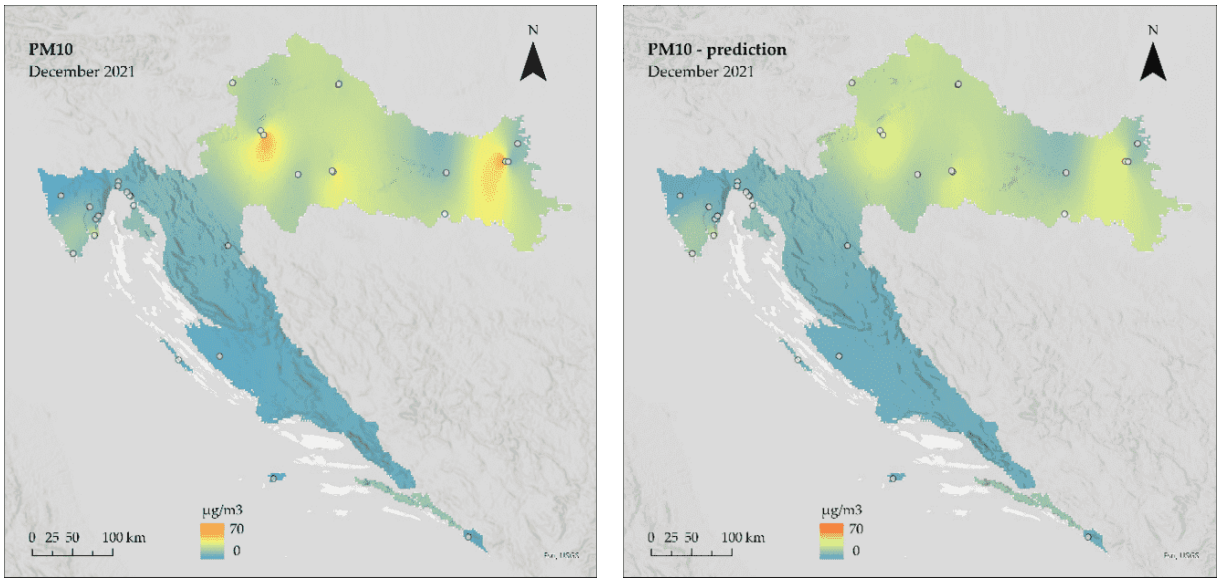
(d)



(e) (f)

Fig. 10. Interpolated monthly PM_{10} concentrations: (a) June from in-situ data (b) June from remote sensing data (c) July from in-situ data (d) July from remote sensing data (e) August from in-situ data (f) August from remote sensing data
Source: own study

The distribution of PM_{10} in months of non-heating season clearly shows us three hotspots (even four for June), all in the northern part of the country, but distributed longitudinally (from west to the east). The central hotspot appears around the urban area of the capital city, Zagreb, and the one in the far east around the urban area of the city of Osijek. Developed models again showed a good performance with ability to find mentioned hotspots.



(a) (b)

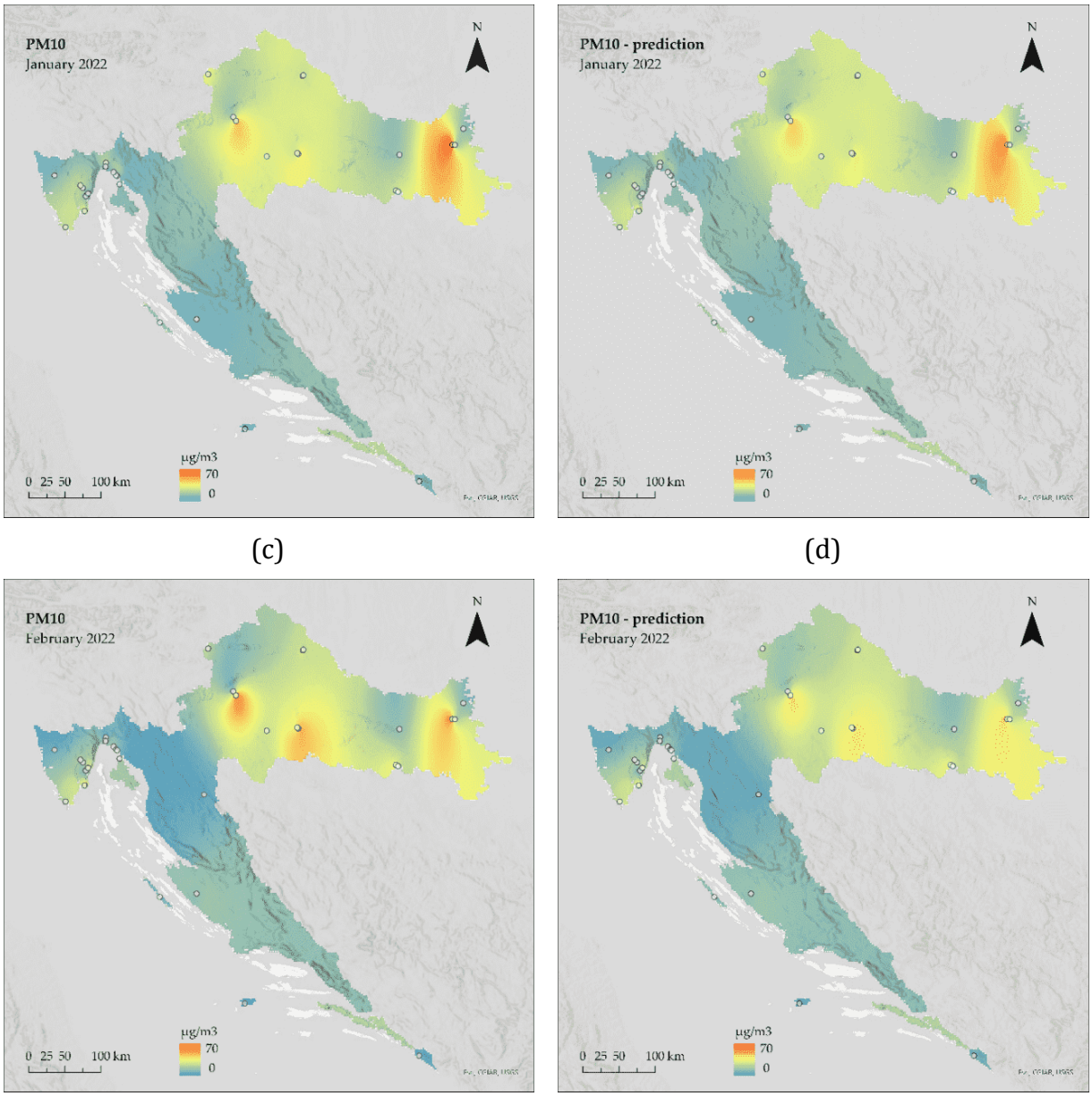


Fig. 11. Interpolated monthly PM_{10} concentrations: (a) December from in-situ data (b) December from remote sensing data (c) January from in-situ data (d) January from remote sensing data (e) February from in-situ data (f) February from remote sensing data Source: own study

When looking at the PM_{10} values in months of the heating season, and those of the non-heating, there is not so much change, even though it was expected that the values during heating season will be much higher. Several studies (Wang et al., 2021; Kumar et al., 2022) connected high PM_{10} emissions to sandstorms and dry climate, neither of which are common in Croatia. There are three hotspots found in the continental part of the country, two of them again located around the urban areas of the cities Zagreb and Osijek, and the third, between, in the town of Kutina.

A common thing for all developed monthly models is that the predicted values slightly underestimate the actual ones and appear slightly lower. However, all models have shown the general ability to estimate $PM_{2.5}$ and PM_{10} levels, even in areas without high pollution. Furthermore, all models can effectively detect all $PM_{2.5}$ and PM_{10} hotspots.

Conclusion

This study proposed a new approach to estimate ambient concentrations of $PM_{2.5}$ and PM_{10} from TROPOMI and other open-source remote sensing data available in GEE.

On a monthly time scale, the Random Forest machine learning method was successfully used to create $PM_{2.5}$ and PM_{10} models for the Republic of Croatia. All monthly models show a moderate to high correlation between in-situ and estimated $PM_{2.5}$ and PM_{10} values, with overall better results for $PM_{2.5}$ than for PM_{10} concentrations. Regarding $PM_{2.5}$ models, the model with the highest correlation ($r = 0.78$) is for January. The PM_{10} model with the highest correlation ($r = 0.79$) is for December.

The spatial distribution of $PM_{2.5}$ concentrations for months during the heating season revealed significant variations in $PM_{2.5}$ pollution in the continental part of Croatia, which motivates the development of regional models and opens a space for new research. Observed seasonal changes could be linked to heating emissions and multiple geographical and meteorological conditions. Furthermore, the next step would be to create seasonal models using the same methodology proposed by this research. The advantage of this approach is that it combines multiple parameters from different sources in order to answer the challenges in the composition of observed air pollutants and that it can be easily adopted in new study areas. Limitations are presented in the form of missing data due to weather conditions.

A common thing for all developed monthly models is that the predicted values underestimate the actual ones and appear slightly lower. However, all models have shown the general ability to estimate $PM_{2.5}$ and PM_{10} levels, even in areas without high pollution. Furthermore, all models can effectively detect all $PM_{2.5}$ and PM_{10} hotspots.

References

Air Pollution, URL 1: https://www.who.int/health-topics/air-pollution#tab=tab_1 [27.10.2022].

Anderson J.O., Thundiyil J.G., Stolbach A. (2012). Clearing the air: a review of the effects of particulate matter air pollution on human health. Journal of medical toxicology, 8(2), 166–175.

Borsdorff T., Hu H., Hasekamp O., Sussmann R., Rettinger M., Hase F., ... & Landgraf, J. (2018). Mapping carbon monoxide pollution from space down to city scales with daily global coverage. Atmospheric Measurement Techniques, vol. 11, no. 10, pp. 5507–5518.

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.

Clemente Á., Yubero E., Nicolás J. F., Caballero S., Crespo J., Galindo N. (2022). Changes in the concentration and composition of urban aerosols during the COVID-19 lockdown. Environmental Research, 203, 111788.

Farr T.G., Rosen P.A., Caro E., Crippen R., Duren R., Hensley S., Kobrick M., Paller M., Rodriguez E., Roth L., Seal D., Shaffer S., Shimada J., Umland J., Werner M., Oskin M., Burbank D., and Alsdorf, D.E. (2007). The shuttle radar topography mission: Reviews of Geophysics, vol. 45, no. 2, RG2004, https://doi.org/10.1029/2005RG000183).

Gautam S., Elizabeth J., Gautam A. S., Singh K., Abhilash P. (2022). Impact Assessment of Aerosol Optical Depth on Rainfall in Indian Rural Areas. Aerosol Science and Engineering, 1–11.

GFS: Global Forecast System 384-Hour Predicted Atmosphere Data, URL 3: https://developers.google.com/earth-engine/datasets/catalog/NOAA_GFS0P25#description [27.10.2022].

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) [Data set].

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.

Kelly F.J., Fussell J.C. (2012). Size, source and chemical composition as determinants of toxicity attributable to ambient particulate matter. Atmospheric environment, 60, 504–526.

Kim K.H., Kabir E., Kabir S. (2015). A review on the human health impact of airborne particulate matter. Environment international, 74, 136–143.

Kleipool Q., Ludewig A., Babić L., Bartstra R., Braak R., Dierssen W., ... & Veefkind P. (2018). Pre-launch calibration results of the TROPOMI payload on-board the Sentinel-5 Precursor satellite. Atmospheric Measurement Techniques, vol. 11, no. 12, pp. 6439–6479.

Kumar R.P., Perumpully S.J., Samuel C., Gautam S. (2022). Exposure and health: A progress update by evaluation and scientometric analysis. Stochastic Environmental Research and Risk Assessment, 1–13.

Kumar R.P., Samuel C., Gautam S. (2022). A bibliometric and scientometric: analysis towards global pattern and trends related to aerosol and precipitation studies from 2002 to 2022. Air Quality, Atmosphere & Health, 1–16.

Kvaliteta zraka u Republici Hrvatskoj (*Air quality in the Republic of Croatia*), URL 2: http://iszz.azo.hr/iskzl/index.html [27.10.2022].

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 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.

Pal M. (2005). Random forest classifier for remote sensing classification. International journal of remote sensing, 26(1), 217–222.

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.

Pope III C.A., Dockery D.W. (2006). Health effects of fine particulate air pollution: lines that connect. Journal of the air & waste management association, 56(6), 709–742.

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.

Singh R., Singh V., Gautam A.S., Gautam S., Sharma M., Soni P. S., ... & Gautam A. (2022). Temporal and Spatial Variations of Satellite-Based Aerosol Optical Depths, Angstrom Exponent, Single Scattering Albedo, and Ultraviolet-Aerosol Index over Five Polluted and Less-Polluted Cities of Northern India: Impact of Urbanization and Climate Change. Aerosol Science and Engineering, 1–19.

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.

Understanding Random Forest, URL 4: https://towardsdatascience.com/understanding-random-forest-58381e0602d2 [27.10.2022].

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.

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.