

POLITECNICO DI MILANO

School of Civil, Environmental and Land Management
Engineering - MSc in Civil Engineering for Risk Mitigation



**Air Quality Monitoring and Prediction in Ukraine During
War Crisis Using Satellite Imagery and Artificial Intelligence**

Supervisor: Dr. Marco Scaioni
Co-supervisor: Dr. Mattia Previtali

Thesis by:
Mohammad Mehrabi
Matr. 966959

Academic Year 2022/2023

ACKNOWLEDGEMENT

I would like to express my profound gratitude to my supervisor Dr. Marco Scaioni from the Department of Architecture, Built Environment and Construction Engineering at Politecnico di Milano. It was only thanks to his valuable guidance, constant encouragement, and precise teachings that I could successfully accomplish this thesis. I am also grateful to my co-supervisor Dr. Mattia Previtali for his friendly advice regarding the used software.

Lastly, the author would like to acknowledge Solcast Company (<https://solcast.com>) and PurpleAir Community (<https://www2.purpleair.com>) for providing free and high-quality data that was used during the second phase of this work.

ENGLISH ABSTRACT

Air quality assessment is an important task, due to the adverse effects of air pollution on human health. This importance is more highlighted when it comes to exceptional events such as war. Early 2022 witnessed the start of a military conflict between Ukraine and Russia. As with any similar event, this war influences the environment from different aspects. The objective of this thesis is to assess and predict war-affected air quality in Ukraine, for which, the project draws on two phases:

(i) Phase 1 deals with air quality monitoring (AQM) using Sentinel-5P imagery and google earth engine. The monthly emission of four gaseous pollutants including ozone (O_3), nitrogen dioxide (NO_2), formaldehyde (HCHO), and carbon monoxide (CO) in 2022 is compared to 2019 and 2021 as business-as-usual (BAU) period. The focus of this evaluation is on five major cities, namely Kiev, Kharkiv, Donetsk, Kherson, and Lviv. According to the results, the general trend of the O_3 concentration is found to be increasing, whereas NO_2 , HCHO, and CO follow mostly a decreasing trend. However, the records of Lviv in terms of NO_2 and HCHO indicate significant discrepancies with others cities that are deemed war fronts. Moreover, based on the applied t-test, the differences between the emission in 2022 vs. BAU years in most cases are statistically significant. All in all, this phase revealed evident effects of the ongoing war on the anthropogenic activities in Ukraine, and consequently, changes in air pollution.

(ii) Phase 2 employs a novel optimized machine learning model for predicting war-affected air pollution in Kiev. In so doing, a well-known machine learning model, namely multi-layer perceptron neural network (MLPNN) is coupled with electromagnetic field optimization (EFO) algorithm to predict the daily concentration of particulate matter 2.5 ($PM_{2.5}$). Initially, a dataset is prepared by collecting seven meteorological factors (humidity (H), temperature (T), dew point (DP), wind speed (WSp), wind direction (WD), solar irradiance (SI), sea level pressure (SLP)), three atmospheric factors (the concentration of O_3 , NO_2 , and CO), and one temporal factor (day number (DN)) from different resources. Next, principal component analysis (PCA) is used to determine the most contributive factors, and based on the results, a reduced dataset is created composed of H, T, O_3 , CO, SI, and DN. Four scenarios are defined by considering the reduced/original dataset, along with, predicting the current day/one-day-ahead $PM_{2.5}$. A sensitivity analysis revealed that the most accurate results are achieved for predicting one-day-ahead $PM_{2.5}$ using the reduced dataset. After adjusting the configurations of both MLPNN and EFO, the EFO-MLPNN hybrid is created and its performance is compared to classical MLPNN and another machine learning model called adaptive neuro-fuzzy inference system (ANFIS). According to the prediction results, the EFO-MLPNN with root mean square error (RMSE) $6.68 \mu\text{g m}^{-3}$ and Pearson correlation coefficient (R_P) 0.82 outperformed both MLPNN (RMSE = $7.51 \mu\text{g m}^{-3}$ and $R_P = 0.79$) and ANFIS (RMSE = $7.98 \mu\text{g m}^{-3}$ and $R_P = 0.73$). These findings infer that optimizing the MLPNN by EFO improves prediction accuracy. Hence, the proposed hybrid model is recommended for more practical air quality estimations. Lastly, a monolithic neural-based formula is extracted from the EFO-MLPNN hybrid for the explicit prediction of $PM_{2.5}$.

Notwithstanding a few limitations, this project demonstrates the applicability of Sentinel-5P imagery and machine learning for reliable air quality assessment during the war.

ITALIAN ABSTRACT

La valutazione della qualità dell'aria è importante a causa degli effetti negativi dell'inquinamento atmosferico sulla salute umana. Questa rilevanza è maggiormente evidenziata quando si tratta di eventi eccezionali come quelli bellici. L'inizio del 2022 ha visto lo scoppio di un conflitto militare tra Ucraina e Russia. Come per qualsiasi evento di questo tipo, questa guerra sta avendo influenza sull'ambiente. L'obiettivo di questa tesi è valutare e prevedere la qualità dell'aria a seguito della guerra in Ucraina. La ricerca si è sviluppata in due fasi:

(i) La Fase 1 si occupa del monitoraggio della qualità dell'aria (AQM) utilizzando le immagini Sentinel-5P e la piattaforma Google Earth Engine (GEE). L'emissione mensile di quattro inquinanti gassosi tra cui ozono (O_3), biossido di azoto (NO_2), formaldeide (HCHO) e monossido di carbonio (CO) nel 2022 viene confrontata con il 2019 e il 2021 come periodo "normale". Il focus di questa valutazione è su cinque grandi città: Kiev, Kharkiv, Donetsk, Kherson e Lviv. Secondo i risultati, la tendenza generale della concentrazione di O_3 risulta essere crescente, mentre NO_2 , HCHO e CO presentano sostanzialmente una tendenza decrescente. Tuttavia, i dati relativi alla città di Lviv in termini di NO_2 e HCHO indicano discrepanze significative con altre città colpite più direttamente dagli eventi bellici. Inoltre, sulla base del test t di *Student* applicato, le differenze tra l'emissione nel 2022 e gli anni "normali" nella maggior parte dei casi sono statisticamente significative. Tutto sommato, questa fase ha rivelato evidenti effetti della guerra in corso sulle attività antropiche in Ucraina e, di conseguenza, cambiamenti nell'inquinamento atmosferico.

(ii) La Fase 2 utilizza un nuovo modello di apprendimento automatico ottimizzato per prevedere l'inquinamento atmosferico causato dalla guerra nella città di Kiev. In tal modo, un noto modello di apprendimento automatico vale a dire la rete neurale MLPNN è accoppiato con l'algoritmo di ottimizzazione del campo elettromagnetico (EFO) per prevedere la concentrazione giornaliera di particolato 2.5 ($PM_{2.5}$). Inizialmente, viene predisposto un set di dati raccogliendo sette fattori meteorologici (umidità (H), temperatura (T), punto di rugiada (DP), velocità del vento (WSp), direzione del vento (WD), irradianza solare (SI), pressione al livello del mare (SLP)), tre fattori atmosferici (la concentrazione di O_3 , NO_2 e CO) e un fattore temporale (numero del giorno (DN)). Successivamente, l'analisi delle componenti principali viene utilizzata per determinare i fattori di maggiore contributo e, sulla base ai risultati, viene creato un set di dati ridotto composto da H, T, O_3 , CO, SI e DN. Vengono definiti quattro scenari considerando il set di dati ridotto/originale, insieme alla previsione del giorno corrente sulla base della concentrazione $PM_{2.5}$ del giorno precedente. Un'analisi di sensibilità ha rivelato che i risultati più accurati si ottengono per prevedere il $PM_{2.5}$ con un giorno di anticipo utilizzando il set di dati ridotto. Dopo aver regolato le configurazioni sia di MLPNN che di EFO, viene creato l'ibrido EFO-MLPNN e le sue prestazioni vengono confrontate con il classico MLPNN e un altro modello di apprendimento automatico chiamato sistema adattivo di inferenza neuro-fuzzy (ANFIS). Secondo i risultati della previsione, l'EFO-MLPNN con errore quadratico medio (RMSE) $6,68 \mu g m^{-3}$ e coefficiente di correlazione di Pearson (R_P) 0,82 ha superato sia MLPNN (RMSE = $7,51 \mu g m^{-3}$ e $R_P = 0,79$) che ANFIS (RMSE = $7,98 \mu g m^{-3}$ e $R_P = 0,73$). Questi risultati deducono che l'ottimizzazione dell'MLPNN da parte di EFO migliora l'accuratezza della previsione. Pertanto, il modello ibrido proposto è raccomandato per stime più pratiche della qualità dell'aria. Infine, dall'ibrido EFO-MLPNN viene estratta una formula monolitica su base neurale per la previsione esplicita del $PM_{2.5}$.

Nonostante alcune limitazioni, questo progetto dimostra l'applicabilità delle immagini Sentinel-5P e dell'apprendimento automatico per una valutazione affidabile della qualità dell'aria durante un evento bellico.

INDEX

1. INTRODUCTION.....	1
1.1 Background.....	1
1.2 Literature Review.....	1
1.2.1 Remote Sensing Aspect.....	1
1.2.2 Machine Learning Aspect	4
1.3 Scope and Objectives	6
1.4 Structure of Thesis	7
2. STUDY AREA AND EVENT	7
2.1 Study Area.....	7
2.2 War Event.....	8
3. PHASE 1: AIR QUALITY MONITORING USING SENTINEL-5P.....	9
3.1 Materials and Methods.....	9
3.1.1 Sentinel-5P	9
3.1.2 Data Acquisition Using Google Earth Engine.....	10
3.1.3 Timing and Sampling.....	11
3.1.4 Considered Pollutants	12
3.1.4.1 O ₃	13
3.1.4.2 NO ₂	13
3.1.4.3 HCHO	13
3.1.4.4 CO	14
3.1.4.5 Data Configuration.....	14
3.2 Results and Discussion.....	16
3.2.1 Pollution Observations.....	16
3.2.1.1 O ₃	16
3.2.1.2 NO ₂	21
3.2.1.3 HCHO	26
3.2.1.4 CO	31
3.2.2 Air Quality at the Beginning of the War.....	36
3.2.3 Statistical Analysis and Discussion	39
3.2.4 Further Discussion	43
4. PHASE 2: AIR QUALITY PREDICTION USING MACHINE LEARNING.....	44
4.1 Materials and Methods.....	44
4.1.1 Data Provision and Analysis	44
4.1.1.1 Parameter Selection.....	44
4.1.1.2 Sources.....	45
4.1.1.3 Data Preprocessing and Splitting	46
4.1.1.4 Description and Statistics	47
4.1.2 Artificial Intelligence	50

4.1.2.1	Artificial Neural Network (ANN).....	51
4.1.2.2	Electromagnetic Field Optimization (EFO).....	52
4.1.3	Accuracy Evaluation Method.....	53
4.1.4	Scenarios.....	54
4.2	Results and Discussion.....	54
4.2.1	Feature Selection Using Principal Component Analysis (PCA).....	54
4.2.2	Method Configuration and Scenario Evaluation.....	57
4.2.3	Optimization and Improvement.....	61
4.2.4	Prediction Results and Discussion.....	65
4.2.4.1	Training and Testing Accuracy.....	65
4.2.4.2	Comparison.....	70
4.2.5	A Predictive Formula.....	71
4.2.6	Further Discussion.....	72
5.	CONCLUSIONS.....	74
6.	BIBLIOGRAPHY.....	76

LIST OF FIGURES

Figure 1: Sentinel-5P launched by the European Space Agency (ESA).....	2
Figure 2: The location of Ukraine and its neighbors	8
Figure 3: A view of the ISW interactive map illustrating conflict types and their intensity	9
Figure 4: A view of the GEE tool in the browser.....	11
Figure 5: The location of the selected cities	12
Figure 6: The concentration of O ₃ in (a) 2019, (b) 2021, and (c) 2022	19
Figure 7: Annual trends and box plots of the O ₃ concentration in (a and b) Kiev, (c and d) Kharkiv, (e and f) Donetsk, (g and h) Kherson, and (i and j) Lviv.....	21
Figure 8: The concentration of NO ₂ in (a) 2019, (b) 2021, and (c) 2022.....	24
Figure 9: Annual trends and box plots of the NO ₂ concentration in (a and b) Kiev, (c and d) Kharkiv, (e and f) Donetsk, (g and h) Kherson, and (i and j) Lviv.....	26
Figure 10: The concentration of HCHO in (a) 2019, (b) 2021, and (c) 2022.....	29
Figure 11: Annual trends and box plots of the HCHO concentration in (a and b) Kiev, (c and d) Kharkiv, (e and f) Donetsk, (g and h) Kherson, and (i and j) Lviv.....	31
Figure 12: The concentration of CO in (a) 2019, (b) 2021, and (c) 2022	34
Figure 13: Annual trends and box plots of the CO concentration in (a and b) Kiev, (c and d) Kharkiv, (e and f) Donetsk, (g and h) Kherson, and (i and j) Lviv.....	36
Figure 14: A view of the TROPOMI Explorer webpage.....	37
Figure 15: Changes in the pollutant emissions (a) O ₃ , (b) NO ₂ , (c) HCHO, and (d) CO in the first week of the war	39
Figure 16: A view of the PurpleAir website showing registered ground stations near Kiev.....	46
Figure 17: Variation of the dataset parameters with time	49
Figure 18: The topology of a general MLPNN(n, m, z)	51
Figure 19: Three polarities and forces among the EPs	52
Figure 20: Scree plot of the PCA analysis.....	55
Figure 21: Topology of the used MLPNN(6, 8, 1) in Scenario 4	62
Figure 22: Hybridization of the MLPNN using EFO	63
Figure 23: Optimization curves for the tried S _{PopS} of the EFO algorithm.....	64
Figure 24: Sensitivity of training and testing RMSEs to the S _{Pop}	64
Figure 25: Target and predicted patterns of PM _{2.5} in training phase of (a) ANFIS(7), (b) MLPNN(6, 8, 1), and (c) EFO(95)-MLPNN(6, 8, 1).....	67
Figure 26: Histogram of error and correlation plots in the training phase of (a and b) ANFIS(7), (c and d) MLPNN(6, 8, 1), and (e and f) EFO(95)-MLPNN(6, 8, 1)	68
Figure 27: Target and predicted patterns of PM _{2.5} in testing phase of (a) ANFIS(7), (b) MLPNN(6, 8, 1), and (c) EFO(95)-MLPNN(6, 8, 1).....	69
Figure 28: Histogram of error (a, c, e) and correlation plots (b, d, f) in the testing phase of (a and b) ANFIS(7), (c and d) MLPNN(6, 8, 1), and (e and f) EFO(95)-MLPNN(6, 8, 1).....	70
Figure 29: Comparing the target and predicted PM _{2.5} trends during the studied period	73

LIST OF TABLES

Table 1: Geographical description of the selected sites.....	12
Table 2: Characteristics of the considered pollutants in Earth Engine Data	15
Table 3: Yearly correlation results (R_p) of O_3 for the considered cities	21
Table 4: Yearly correlation results (R_p) of NO_2 for the considered cities.....	26
Table 5: Yearly correlation results (R_p) of HCHO for the considered cities.....	31
Table 6: Yearly correlation results (R_p) of CO for the considered cities.....	36
Table 7: Analysis of pollutant concentrations and changes	40
Table 8: T-test results (p-values) for comparing the pollutant concentrations in 2022 vs BAU years	42
Table 9: Description of the used parameters.....	46
Table 10: Descriptive statistics of the dataset	50
Table 11: Correlation analysis of the parameters.....	50
Table 12: Detailed variance analysis of the PCs	56
Table 13: Rotated factor loadings and communalities for the used rotation methods	56
Table 14: Fixed characteristics of the applied MLPNNs.....	57
Table 15: The results of trial-and-error process for finding the best m in Scenario 1.....	59
Table 16: The results of trial-and-error process for finding the best m in Scenario 2.....	59
Table 17: The results of trial-and-error process for finding the best m in Scenario 3.....	60
Table 18: The results of trial-and-error process for finding the best m in Scenario 4.....	60
Table 19: The accuracy results obtained for four implemented scenarios	61
Table 20: Parameters of the EFO algorithm	65
Table 21: Calculated accuracy criteria for training and testing phases.....	70

1. INTRODUCTION

1.1 Background

Air pollution is among the most serious problems in the world, due to the rapid population growth and its consequences. Activities such as urbanization, motorization, energy production, transportation, heavy industry, etc. are the main causes of air pollutants. As is known, pollution can adversely impact the environment and human health. Therefore, it is considered a concern of not only experts and ecologists, but also ordinary people. According to the estimations by the world health organization (WHO), around 7 million people die annually because of exposure to polluted air, along with nearly 4.2 million premature annual deaths mainly due to health issues including heart disease, pulmonary disease, stroke, children's acute respiratory infections, etc. According to [World Bank \[1\]](#), around 90% of people in low and middle-income countries are dangerously exposed to polluted ambient air. It is even worse for developing countries where the concentration of pollutants and exposure exceed WHO standards. Not surprisingly, the reasons lie in increasing industrial activities which are associated with burning dirty fuels and employing outdated technologies, as well as the lack of essential healthcare awareness and inadequate standards regarding air quality [\[2\]](#).

Ukraine is one of the most critical countries regarding air pollution consequences. WHO reported 14,400 life losses annually because of air pollution in Ukraine [\[3\]](#). Out of 120 countries in 2017, this country had the largest number of fatalities (per every 100,000 people) attributed to atmospheric air pollution [\[4\]](#). Based on another piece of statistics, the rate of deaths caused by this issue is higher than 200 per 100,000.

Despite recent progress carried out for improving air quality assessment in Ukraine (e.g., air quality EU Directives), there are still some limitations (such as insufficient and unfair distribution of ground-based measurement sites [\[5\]](#)) which necessitates using more sophisticated tools like satellite imagery to obtain a comprehensive image of the pollutants over this country.

1.2 Literature Review

1.2.1 Remote Sensing Aspect

In the previous literature, many studies have benefitted from various sophisticated techniques for air quality monitoring (AQM) in different parts of the globe. However, the world has witnessed a paradigm shift in air pollution monitoring from conventional government-operated networks (i.e.,

reference instruments) to more sophisticated mixed networks (i.e., reference-grade monitors and sensor/monitor technologies). These low-cost new technologies received increasing attention later and were recommended in standards of relevant agencies such as the U.S. environmental protection agency (EPA) [6].

Satellite remote sensing (RS) is a highly evolved technique for conducting reliable AQM [7]. Satellite data can be beneficial to estimate emissions, provide evidence for exceptional event declarations, track pollutant plumes, and support air quality forecast results [8]. In this approach, the pollution representatives are sensed through spatial assessment of satellite products using environments such as geographic information system (GIS) [9]. This assessment can also be improved using different sources including additional air pollutant measurements, modeled air pollutant concentrations, and spatial surrogate data such as land-cover characteristics [6].

Sentinel-5 Precursor (a.k.a Sentinel-5P and S5P) (Figure 1) is a recently launched satellite that is able to sense the concentration of various pollutants using four spectrometers with spectral bands in the ultraviolet (UV), visible (UVIS), near-infrared (NIR) and shortwave infrared (SWIR) wavelengths [10, 11]. A review of previous studies indicates successful applications of Sentinel-5P satellite in AQM by detecting changes in pollutants such as sulfur dioxide (SO₂) [12], methane (CH₄) [13], nitrogen dioxide (NO₂) [14], CO [15], formaldehyde (HCHO) [16], ozone (O₃) [17], and aerosols [18].



Figure 1: Sentinel-5P launched by the European Space Agency (ESA) [19]

Up to now, this satellite has provided the primary materials for many AQM studies. For instance, [Ayoobi, Ahmadi \[20\]](#) analyzed how seasonal diversity of energy consumption patterns in the building can affect the concentrations of tropospheric NO₂, CO, and SO₂ in Kabul, Afghanistan. For this purpose, they used Sentinel-5P data which disclosed the decline of the mentioned pollutants going from winter to summer. This reduction was explained by cutting the use of substances like coal, gas, and biomass which are burnt for heating purposes in winter. [Bhatkar, Syamala \[21\]](#) used the NO₂ records of Sentinel-5P to assess the relationship between urbanization and population growth in the United Arab Emirates. Their research showed a meaningful positive correlation between the considered parameters, and hereupon, they recommended further continuous monitoring of this area. The measurements of this satellite have also served other environmental purposes such as mapping important parameters e.g., sun-induced chlorophyll fluorescence [\[22\]](#) and water vapor columns [\[23\]](#).

When it comes to extreme environmental conditions, the effectiveness of satellite-based AQM is more highlighted, due to the limitations in viable ground measurements [\[24\]](#). Scientists have benefitted from Sentinel-5P data to better manage problematic conditions such as post-fire air pollution monitoring (e.g., in Türkiye [\[25\]](#) and Portugal [\[10\]](#)) and detecting Volcanic SO₂ Plumes (in Italy [\[26\]](#) and [\[27\]](#)) in different areas. Recently, the world has witnessed extensive emergency lockdowns, due to the COVID-19 pandemic. In the meantime, many experts could successfully use RS measurements, and particularly Sentinel-5P data, to monitor the effects of these restrictions on air quality all over the globe. For instance, [Sharifi and Felegari \[28\]](#) examined the effect of pandemic quarantine conditions in Tehran, Iran. Utilizing the Sentinel-5P data in 2019 and 2020, they detected that the concentration of NO₂ in 2020 fell by 6%. With the same logic, [Vîrghileanu, Săvulescu \[29\]](#) studied Europe and found an up to 85% decrease in the tropospheric NO₂ in some major cities. It was also revealed that satellite data are in good agreement with ground-based observations (correlation coefficients (R^2) ranged in [0.5, 0.75]). In a study over Tamil Nadu, India, [Muniraj, Panneerselvam \[30\]](#) used the Sentinel-5P coupled with ground measurements to see how air quality is affected by COVID lockdown. As the primary results, they showed significant reductions (i.e., from 50% to 314%) in the emissions of CO, NO₂, SO₂, and O₃ in the period of March-June 2020. Moreover, comparing the satellite acquisitions with in-situ records professed above 83% accuracy for them. Likewise, [Stratoulis and Nuthammachot \[31\]](#) used a combination of Sentinel-5P and ground data to evaluate the influence of the COVID lockdown in Hat Yai, Thailand. In terms of NO₂ and particulate matter i.e., PM₁₀ and PM_{2.5}, their concentrations declined by about 34%, 23%, and 22%, respectively, during the first 3 weeks of quarantine. Further similar efforts can be found adequately in the previous literature [\[32-34\]](#).

Certainly, devastating conflicts that have recently happened/are now happening in different parts of the world (e.g., Afghanistan, Yemen, Syria, etc.) are another type of extreme events that impact the environment in various ways. For example, substantial air pollution effects were recorded for Kuwaiti oil fires in Gulf War [35]. In general, suspension of normally active systems (e.g., industry and transportation), as well as inducing detrimental military activities (e.g., the movements of military equipment, fire, and explosions), can impact the quality of air in different scales [36]. In such cases, satellite observations can play an excellent complementary role for AQM. Reading the literature exposes promising applications of Sentinel-5P imagery for monitoring the pollution caused by military explosion events. For example, [Ali, Abouleish \[37\]](#) focused on the powerful blast of an ammonium nitrate storehouse at Beirut Port on 4 August 2020. Tracing the NO₂ emissions revealed higher concentrations of this pollutant compared to the pre-event period.

1.2.2 Machine Learning Aspect

The advent of machine learning (ML) models has revolutionized various scientific domains, especially when it comes to solving complicated engineering problems. Provided with a valid dataset, an ML model is capable of pattern recognition by considering a number of influential factors and predicting the behavior of the intended parameter(s) [38, 39]. High accuracy of non-linear prediction, excellent flexibility, time efficiency, and inexpensive implementation are among the most important advantages of ML tools that make them popular in various engineering domains including civil [40] and environmental engineering [41].

More particularly for air quality assessments, different ML methodologies have been so far applied to predict atmospheric pollution [42-44]. A review of the past decade's literature denotes a significant evolution of studies that employed ML for air quality prediction (see Figure 2 of [45] and [46]). In this regard, it has been shown that these models are able to accurately predict the presence of pollutants by analyzing temporal data and meteorological factors (e.g., humidity, temperature, and wind effects). Artificial neural networks (ANNs) [47], neuro-fuzzy (NF) [48] systems, support vector machine (SVM) [49], and random forest (RF) [50] are some of the most popular ML models that have attracted the scientists' attention for air pollution predictions. [Maleki, Sorooshian \[51\]](#), for instance, used an ANN to predict air quality index (AQI) and air quality health index (AQHI) for Ahvaz, Iran, which is known as one of the most polluted cities in the world. These indices were calculated based on six pollutants, namely O₃, NO₂, PM₁₀, PM_{2.5}, SO₂, and CO. The model explored the effect of temporal (time and date) and meteorological (wind speed, air pressure, temperature, dew point, and rainfall) parameters, as well as 3 and 6 h ahead pollution concentrations to predict the AQI and AQHI. Due to the nice agreement between the measured and

predicted values, they concluded the suitability of the ANN for air quality prediction. Furthermore, many comparative efforts have been conducted to comparatively evaluate the performance of various ML models. As an example study, [Zhang, Wang \[52\]](#) employed an ML technique called light gradient boosting machine (LightGBM) to predict PM_{2.5} using an integration of temporal, meteorological, statistical, and air quality features. They also showed the outperformance of the proposed LightGBM in comparison with benchmark models, namely Gradient Boosting Decision Tree (GBDT), Extreme Gradient Boosting (XGboost), Catboost, and Deep Neural Network (DNN). Further comparative efforts can be adequately found in relevant literature review efforts [\[53\]](#).

More recent attempts by ML developers have resulted in the introduction of novel hybrid ML techniques that can be used for a wide range of prediction problems. These models are called hybrid because they are composed of two (or more) pivotal algorithms, the first one is a conventional ML such as ANN that represents the basis of the models and the second part is a metaheuristic optimization algorithm that plays the role of an optimizer. Many studies have proven that considering this strategy leads to surmounting computation drawbacks of conventional predictive models (e.g., local minima issue in ANN), and consequently, creating a more powerful model [\[54, 55\]](#). Engineers have also used these hybrid models (i.e., a conventional ML + a metaheuristic algorithm) in the field of air quality forecast [\[56, 57\]](#). Genetic algorithm (GA), differential evolution (DE), and particle swarm optimization (PSO) are among the most famous metaheuristic techniques. [Yonar and Yonar \[58\]](#) employed these three algorithms to optimize a so-called ML model adaptive neuro-fuzzy inference system (ANFIS) for predicting PM_{2.5} in İstanbul, Turkey. This task was fulfilled by taking into account the effect of gaseous pollutants including SO₂, O₃, NO₂, and CO, and meteorological parameters including wind speed, temperature, wind gust, humidity, and pressure. According to their results, the optimized ANFIS could surpass its classical version. Similar applications can be found in earlier works (e.g., multi-verse optimization (MVO) for optimizing long short-term memory (LSTM) model [\[59\]](#), cuckoo search algorithm (CS) and grey wolf optimizer (GWO) for optimizing support vector regression (SVR) model [\[60\]](#)). As far as the ANNs are concerned, experts have proposed optimized versions for air quality prediction. Genetic simulated annealing (GSA) algorithm was coupled with ANN by [Kang and Qu \[61\]](#) for predicting the AQI in Lanzhou, China. Referring to the calculated mean absolute percentage errors 5.51% and 1.08%, respectively for the classical and optimized ANNs, it was deduced that the executed optimization could satisfactorily enhance the accuracy of ANN. Similarly, [Xia \[62\]](#) improved the prediction accuracy of ANN by 4% when it was combined with fruit fly optimization (FFO) algorithm.

1.3 Scope and Objectives

Considering the ongoing war crisis in Ukraine, its air quality has certainly experienced significant changes as a result of shifts in anthropogenic activities that can affect the pollution emission rate. Similar trends have been previously observed in this area, due to the COVID-19 lockdowns in recent years [63, 64]. As with any other major conflict, this event incurred changes in the normal patterns of anthropogenic activities, and also caused detriment activities, which altogether call for conducting proper AQM in order to reach the maximum preparedness in dealing with the current situation.

As elucidated above, previous sections well described different aspects of AQM studies. Apart from the importance of continuous AQM for mitigating air pollution consequences in normal circumstances, the role of this task is more critical in the time of extreme events such as war and natural disasters. Taking a look at the literature demonstrates that, despite its critical importance, very few studies have worked on this subject before. For example, in studies by [Rawtani, Gupta \[65\]](#) and [Pereira, Bašić \[36\]](#) the environmental impacts (air pollution included) of the Russia-Ukraine crisis have been investigated in brief along with economical and social aspects. In a more particular effort, [Zhang, Hu \[66\]](#) used tropospheric monitoring instrument (TROPOMI) and environmental trace gases monitoring instrument (EMI) to evaluate the changes in NO₂ in both Ukrainian and Russian marginal territories from January to July. [Zalakeviciute, Mejia \[67\]](#) studied the air pollution evolution of Ukraine during the first weeks of the war using satellite imagery coupled with land measurement data. They developed maps for several pollutants namely, NO₂, CO, SO₂, O₃, and PM_{2.5}, and calculated quantitative increments and reductions in each one. Notwithstanding the valuable outcomes of these studies, each suffered from limitations. For example, [Zhang, Hu \[66\]](#) focused only on NO_x pollutant, and the findings of [Zalakeviciute, Mejia \[67\]](#) is concerned with the initial weeks of the war. Hence, this thesis is dedicated to conducting proper AQM over the war-affected Ukraine with a focus on bridging previous gaps of knowledge. Utilizing state-of-the-art technologies, it is tried to keep the AQM updated with the latest developments. Hereupon, the thesis is conducted in two major phases:

- (i) Phase 1 uses satellite imagery and statistical analysis to focus on AQM over the year 2022 and compares it with previous years in order to detect noticeable changes in the emission of well-known air pollution indicators.
- (ii) Phase 2 presents a hybrid artificial intelligence technique to predict air pollution using various data sources and yields a mathematical explicit formula for convenient applications.

Besides revealing the unpleasant effects of the current crisis on air quality to the world, the findings of this project might be of interest to the Ukrainian authorities and decision-makers to take appropriate measures in tackling air quality issues, especially in war-affected regions. It also may shed light on future public health studies that deal with the long-terms effects of toxic air pollution.

1.4 Structure of Thesis

Following the introduction, the rest of the thesis is structured as follows: Section 2 introduces the study area and its characteristics, as well as the details of the Russia-Ukraine crisis. Thereafter, two major phases are explained each of which contains specific material and methods followed by the corresponding results and discussion. The last section is the conclusion which reflects the main findings of this research.

2. STUDY AREA AND EVENT

2.1 Study Area

Lied within the longitude $22^{\circ} 12'$ to $40^{\circ} 07'$ E and latitude $44^{\circ} 25'$ to $52^{\circ} 11'$ N, the independent republic of Ukraine is the second largest country in Europe (about 6% of the European subcontinent). Figure 2 shows the relative location of this country. Considering the neighbors, Ukraine shares boundaries with Russia in the east, Russia and Belarus in the north, Poland in the west, and Moldav and Romania in the south.

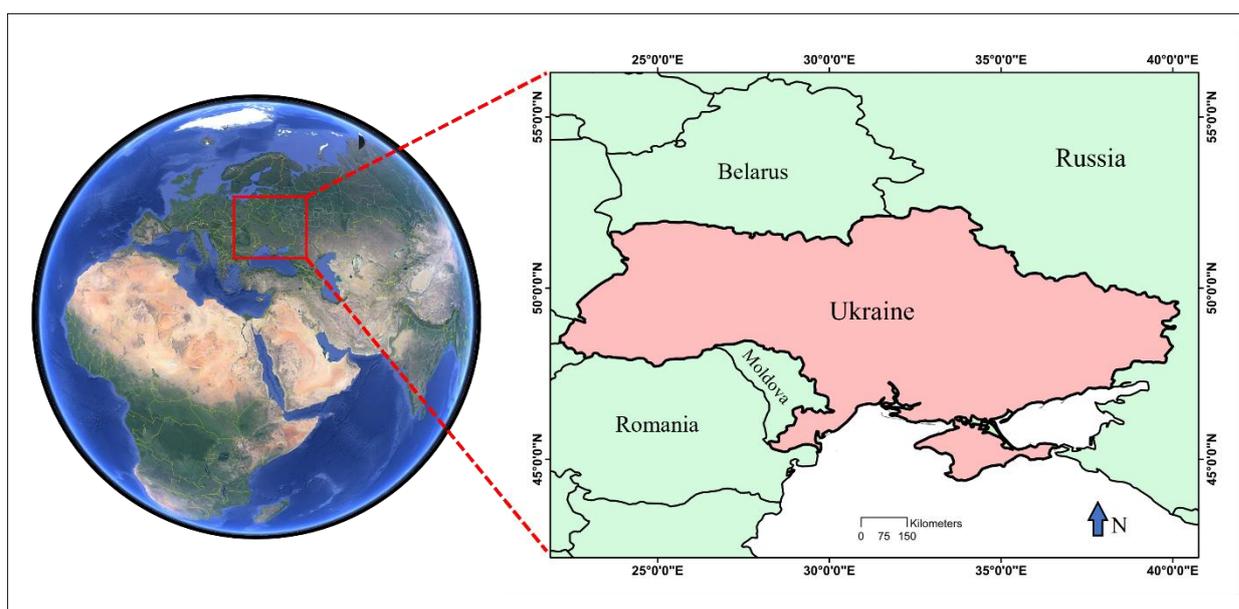


Figure 2: The location of Ukraine and its neighbors

Regarding the climatological characteristics of Ukraine, the temperate climate is dominant. According to meteorological statistics averaged for [1991-2020], the minimum and maximum monthly temperature is -3.35 °C and 21.46 °C in January and July, respectively. Likewise, the monthly precipitation ranges from 35.43 mm and 67.17 mm in February and June, respectively [68]. It is worth mentioning that approximately 16% of the total land is covered by forests with an uneven distribution [69].

The area is roughly 603,550 km² which makes Ukraine the 45th largest country in the world. According to online resources, the population in 2021 has been reported 43,792,855, and two-thirds of them are residents of urban areas [70]. The war has forced more than 10 million Ukrainian people to leave their homes, out of them, nearly 5 million have left Ukraine [71].

2.2 War Event

Rooted in the past political tensions between the two countries, the 2022 Ukraine-Russia war started with some conflict warnings. Following the declaration of "special military operations" on February 21 by Russia, tensions increased, and eventually, on February 24, Russia launched a large-scale invasion toward Ukrainian territories comprising missile and artillery attacks [72].

Numerous attacks and counterattacks by both sides involved bombing and shelling in major cities have led to considerable damage and destruction to infrastructures. According to Statista [73], the number of civilian casualties until January 29, 2023, exceeds 7000 deaths and 11000 injuries. As for military fatalities, BBC [74] reflected the US estimates accounting for 200,000 military casualties on all sides. Apart from the evident loss of lives and physical damage to infrastructure and vital supplies, this war brings massive shrinkage for Ukraine's economy (predicted up to 45% by World Bank [75]).

Online resources such as the institute for the study of war (ISW) [76] provide well-updated interactive maps regarding occupied territories and the advances of Russian and Ukrainian forces. Until January 20, 2023, the number of total conflict events exceeds 35000. As shown in Figure 3, this number comprises battles, explosions and remote violence, riots, and violence against civilians [77].

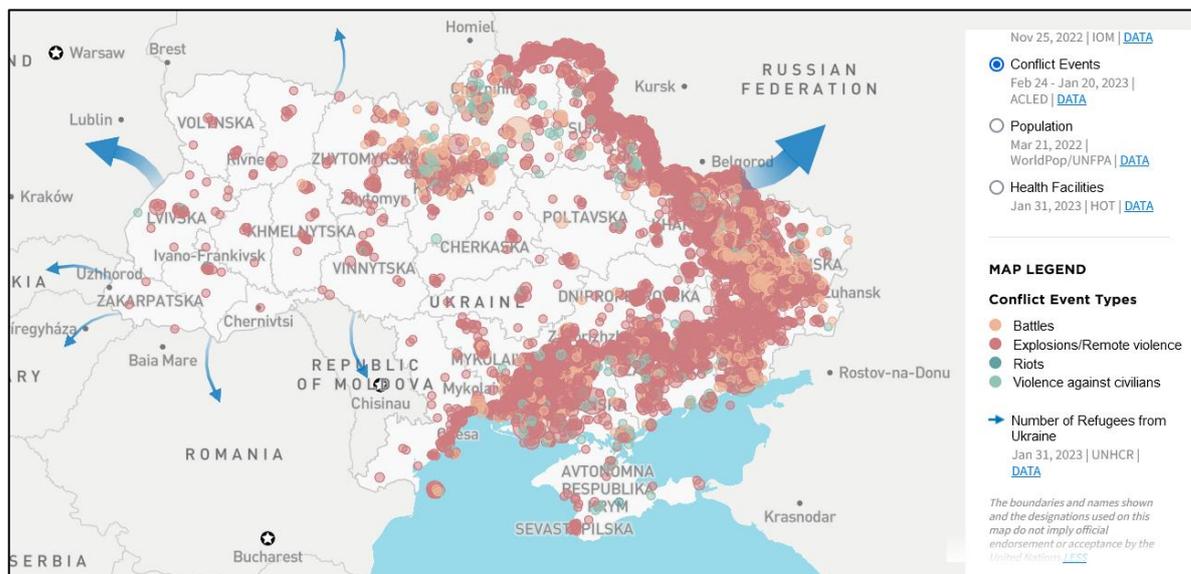


Figure 3: A view of the ISW interactive map illustrating conflict types and their intensity [77].

Employing heavy military equipment in this war has resulted in evident causes of air pollution. Besides the fires and explosions, the movements of military equipment and troops need to be regarded [36]. In the current crisis, a large number of heavy military equipment are being used by both sides [78-80]. In particular, they consist of ground vehicles (e.g., tanks and armored vehicles), aircraft (e.g., bombers), and heavy artillery [81].

3. PHASE 1: AIR QUALITY MONITORING USING SENTINEL-5P

As explained earlier, the first phase of this study is concerned with satellite-based AQM. For this purpose, the concentrations of air pollutants during the war are compared to past records in order to monitor the changes. These illustrative assessments are further accompanied by quantitative statistical evaluations.

This section is composed of two sub-sections describing, first, the materials and methods used for Phase 1, followed by presenting the results and relevant discussion.

3.1 Materials and Methods

3.1.1 Sentinel-5P

Launched on 13 October 2017, Sentinel-5P is the first Copernicus mission dedicated to AQM. This mission belongs to the European Union's Copernicus Earth program. The satellite is capable of

addressing the geographic distribution of pollutants, owing to the data collection at high altitudes (i.e., 824 km from the earth) [82]. Sentinel-5P flies in a Sun-synchronous orbit and its local overpass time is 13:30. TROPOMI is the single instrument carried by Sentinel-5P. With a 2600 km wide swath, TROPOMI presents a near-daily global coverage to collect daily data from the atmosphere of the Earth in several bands including UV, UVIS, NIR, and SWNIR [11, 13]. The spatial resolutions, in the beginning, were approximately $3.5 \times 7 \text{ km}^2$ for the UV and visible bands, and $7 \times 7 \text{ km}^2$ for the SWNIR. After improvements carried out on August 2019, these resolutions reached $3.5 \times 5.5 \text{ km}^2$ and $7 \times 5.5 \text{ km}^2$, respectively [32]. Three processing types are considered for Sentinel-5P data: (i) near real-time (NIR), (ii) offline (OFFL), (iii) and reprocessing. Preparing the data of OFFL and reprocessing types takes a long time (i.e., between 12 hours and 5 days after sensing), while the data of NRT is ready within 3 hours after sensing [83]. Further details regarding the Sentinel-5P mission are available on the ESA website [84].

The observations of Sentinel-5P are available mainly through Copernicus Open Access Hub [85] which guides to Sentinel-5P Pre-Operations Data Hub [86]. However, there are some other portals such as Sentinel Hub EO Browser [87] and Earth Engine Apps TROPOMI Explorer [88] that provide different formats of this data.

3.1.2 Data Acquisition Using Google Earth Engine

In order to acquire data for this project, the google earth engine (GEE) is used to extract Sentinel-5P image collections and map the results. As a well-known cloud computing platform, Google launched the GEE in 2010 [89]. By providing cloud computation, this tool has shown high efficacy in dealing with various geospatial problems with big data analysis. An advantage of the GEE in comparison with similar cloud-based platforms is supporting a wider range of geospatial data such as early Landsat, MODIS, and Sentinel missions. Giving free access to the users is another notable merit of this platform, especially for researchers in less developed countries [90].

The users of the GEE can find publicly downloadable earth observation data on large scales (i.e., petabytes) stored in big repositories of raw remotely sensed images which also contain preprocessed, cloud-removed, and mosaicked imagery [91]. Thanks to sophisticated algorithms and packages (e.g., image processing, ML, chart developers, etc.), both expert and non-expert users are enabled to analyze geographic data. These properties altogether have made the GEE the most popular cloud computing platform in earth system science [90]. So far, the GEE has assisted engineers to analyze environmental data related to changes in climate [92], forest [93], urbanization [94], land use [95], air quality [96], etc. This is worth noting that by utilizing the timelapse tool, one is capable of monitoring and tracking environmental changes, even in the past few decades.

Leveraging Google’s computational infrastructure, the GEE enables the user to carry out time-efficient parallel data processing. As far as the working environment is concerned, a user-friendly code editor, i.e., a web-based Integrated Development Environment (IDE), is designed for writing and implementing complex scripts using JavaScript API [97]. Also, the explorer is a straightforward web app interactor for performing simple tasks such as exploring data catalog and visualization. Figure 4 illustrates the GEE environment that is used in this research.

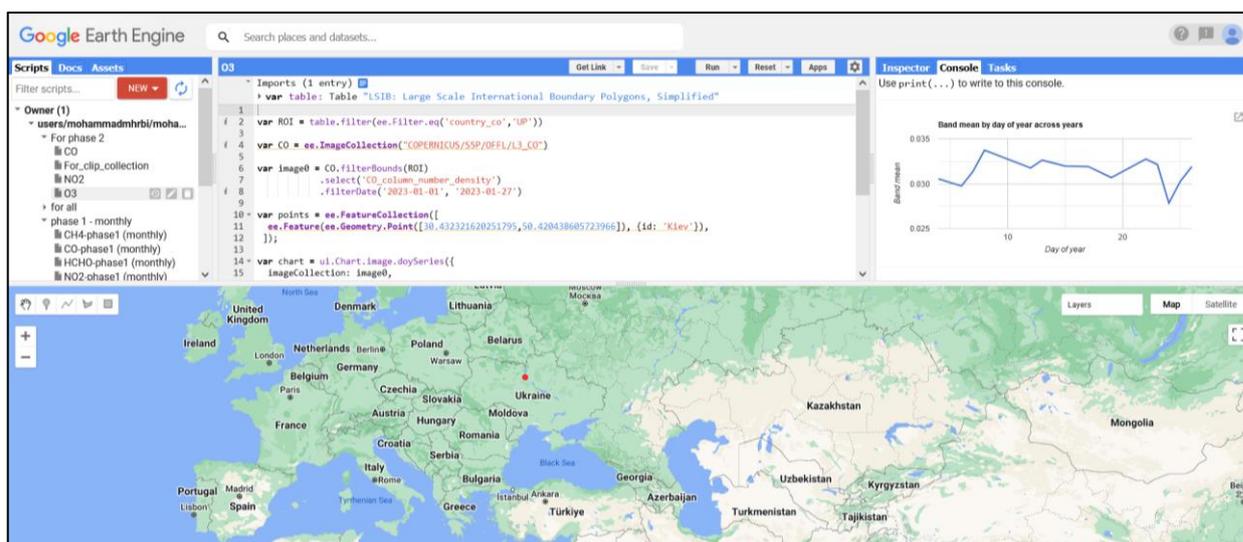


Figure 4: A view of the GEE tool in the browser.

3.1.3 Timing and Sampling

For such projects, it is important to have proper timing strategies to achieve the highest reliability of the results. The concentration of pollutants during 2022 represents the war-affected records, while similar records belonging to 2019 and 2021 are deemed as business-as-usual (BAU) data. Not surprisingly, the reason for dropping 2020 data is the lockdown measures considered for the COVID-19 pandemic in that year (the Ukraine government fixed sanitary emergency as of 20 March 2020) [67]. Due to the residual pandemic restriction in 2021, however, alleviated with respect to 2020, the records of 2019 represent the most BAU condition in this study.

For each month of 2022, daily records of an air pollutant are averaged, and a monthly map is accordingly generated for the whole country. This map is then compared with the corresponding maps pertaining to 2019 and 2021 to visualize the differences. Apart from month-to-month comparison, the yearly trends of increments and reductions are compared. Since a single value cannot well represent air pollution in a large country like Ukraine, the concentrations are quantitatively reported and assessed for five major cities, namely Kiev, Kharkiv, Donetsk, Kherson,

and Lviv. The location and coordinates of these cities are presented in Figure 5 and Table 1. As is shown, the first four sites are on marginal fronts that are highly engaged with the war, while Lviv can be treated as a less critical city. As a benefit of this distinction, it allows for underlining the effect of geographical locations on the pollution potential of a place.

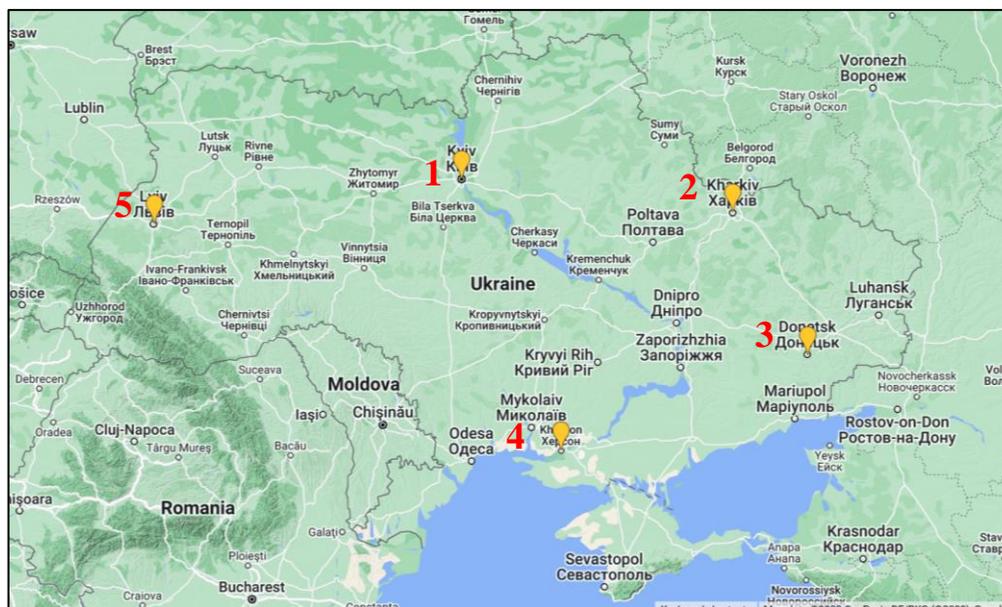


Figure 5: The location of the selected cities

Table 1: Geographical description of the selected sites

#	Place	Longitude	Latitude
1	Kiev	30.523333	50.450001
2	Kharkiv	36.232845	49.988358
3	Donetsk	37.805279	48.002777
4	Kherson	32.616867	46.635417
5	Lviv	24.031111	49.842957

3.1.4 Considered Pollutants

The concentration and composition of pollutants in the outdoor air determine the severity of their adverse health effects [98]. Among several pollutants in cities' outdoor air, O₃, SO₂, carbon monoxide (CO), and nitrogen oxides (NO_x) are among the most significant ones [99].

As explained in earlier sections, Sentinel-5P senses the concentration of six air pollutants, namely SO₂, CO, NO₂, O₃, CH₄, and HCHO. Descriptive details regarding these parameters can be found in Earth Engine Data Catalog-Sentinel-5P [100].

For the case of this research, initially, it was first anticipated to evaluate all six parameters, but due to data collection issues encountered for the SO₂ and CH₄, the observations are limited to O₃, NO₂, HCHO, and CO. The mentioned issues can be described as (i) significant data gaps and (ii) the

presence of negative concentration values (i.e., negative vertical column densities). According to the Mission Status Reports [101], the gaps were due to acquisition faults or satellite/instrument disruption, while negative values of SO₂ and CH₄ were due to noise on the data and reflectance problems. Detailed status of the Sentinel-5P mission including a list of the major data gaps is presented by ESA [101]. Furthermore, complementary information for each pollutant can be found in TROPOMI Mission Performance Centre [102].

In the following, the used pollutants are described.

3.1.4.1 O₃

Existing in both troposphere and stratosphere layers, O₃ is one of the most predominant pollutants. Tropospheric O₃ (a.k.a. ground-level O₃) is a phytotoxic¹ pollutant that is basically generated from photochemical reactions [103]. These reactions involve volatile organic compounds (VOCs) and NO_x. For instance, CH₄ is known as the primary anthropogenic VOC for the production of O₃ in the global troposphere. However, particularly in polluted areas, nonmethane VOCs play a greater role in the formation of O₃ [104]. The uncontrolled increment of O₃ can increase morbidity and mortality in humans. Promoting problems in the respiratory and cardiovascular system (as short-term effects) and respiratory mortality and asthmatics problems (as long-term effects) are examples of these consequences [105].

3.1.4.2 NO₂

As a major trace gas, NO₂ is the product of various anthropogenic activities. Based on the report from European Environment Agency [106], motor vehicle exhaust fumes are the source of above 60% of NO₂ in European cities. Other sources of this pollutant are industrial and manufacturing activities (e.g., oil refinement and electricity generation) and natural causes (e.g., microbiological activities in soils and wildfire). Apart from the respiratory issues in the case of NO₂ inhalation, its interaction with the atmospheric components results in detrimental phenomena such as acid rain [3].

3.1.4.3 HCHO

Characterized by high flammability, colorlessness, and a sharp odor, HCHO is a carcinogen gas that can highly compromise human health [107, 108]. It plays an intermediate role in almost every oxidation chain of non-methane volatile organic compounds (NMVOCs). A property of this gas is

¹ Phytotoxicity refers to adverse impacts on plant growth, physiology, or metabolism caused by chemical substances

a short lifetime in the atmosphere (i.e., a few hours) that provides a good approximation of its sources. The oxidation of CH₄ is considered a primary source of this gas in the remote atmosphere, while on a global scale, it is produced by wildfire and vegetation. As for the anthropogenic origins, traffic and petrochemical industries (e.g., producing fertilizers) can be mentioned [3].

3.1.4.4 CO

CO is a poisonous gas without any color, taste, and odor which is produced in incomplete combustions. A significant role of this pollutant is increasing the lifetime of other greenhouse gases such as CH₄ and tropospheric O₃. Likewise, the combined action of other pollutants is enhanced in places with elevated concentrations of CO. Natural causes (e.g., decomposition of vegetation and animals, volcanic eruptions, and forest fires) and human activities (e.g., fossil fuel consumption, tobacco smoke, and waste incineration) account for nearly 40% and 60% of the CO generation, respectively [109]. This gas causes, in general, toxic effects on the organs of tissues that are characterized by high oxygen consumption suchlike the heart and brain. Also, it particularly incurs damage to pregnant women and their developing fetuses [110].

3.1.4.5 Data Configuration

Data products of Sentinel-5P can be classified into three levels, namely (i) Level-0, (ii) Level-1B, and (iii) Level-2. Omitting Level-0 which contains instrument source packets and is not accessible to end-users, Level-1B represents geo-located and radiometrically corrected top-of-atmosphere radiances which include radiance products, calibration products, and engineering products. Level-2 data is produced by processing Level-1B data and forms the material of this research.

Level-1B and Level-2 products can be downloaded from Sentinel-5P Pre-operations Data Hub [111]. However, in the current project, Sentinel-5P satellite images that reveal the concentration of O₃, NO₂, HCHO, and CO over Ukraine are acquired and processed using the GEE. With reference to Earth Engine Data Catalog - Sentinel-5P [112], Table 2 collects the characteristics of the pollutants (i.e., the unit of measurement, relevant image collection, used band name, and band name description).

Table 2: Characteristics of the considered pollutants in Earth Engine Data

Pollutant	Unit	Image Collection	Band Name	Band Name Description
O ₃	μmol m ⁻²	COPERNICUS/S5P/OFFL/L3_O3	O3_column_number_density	Total atmospheric column of O ₃ between the surface and the top of atmosphere, calculated with the GODfit algorithm [113]
NO ₂	μmol m ⁻²	COPERNICUS/S5P/OFFL/L3_NO2	NO2_column_number_density	Total vertical column of NO ₂ (ratio of the slant column density of NO ₂ and the total air mass factor)
HCHO	μmol m ⁻²	COPERNICUS/S5P/OFFL/L3_HCHO	tropospheric_HCHO_column_number_density	Tropospheric HCHO column number density
CO	μmol m ⁻²	COPERNICUS/S5P/OFFL/L3_CO	CO_column_number_density	Vertically integrated CO column density

In order to create the map of each pollutant, the below steps are taken in the GEE environment:

- (i) Introducing Ukraine as the area of interest (AOI);
- (ii) Introducing Kiev, Kharkiv, Donetsk, Kherson, and Lviv as sample points by importing their coordinates;
- (iii) Calling the image collection of the intended pollutant from Sentinel-5P offline records (see Table 2);
- (iv) Filtering the image collection for (a) the intended band (see Table 2) and (b) the required period, then (c) calculating the monthly mean, and (d) clipping for the AOI;
- (v) Extracting the monthly mean for the five cities.

3.2 Results and Discussion

This section presents and discusses the results of Phase 1. The monthly concentration of the selected four pollutants (i.e., O₃, NO₂, HCHO, and CO) is evaluated in Ukraine, particularly recorded for five cities (i.e., Kiev, Kharkiv, Donetsk, Kherson, and Lviv), in 2022 versus BAU years (i.e., 2019 and 2021). The results are statistically analyzed to highlight the changes as the effects of the ongoing conflict. Moreover, utilizing another type of Sentinel-5P data, the changes in the concentration of the mentioned pollutants are elaborated with a focus on the initial days of the war.

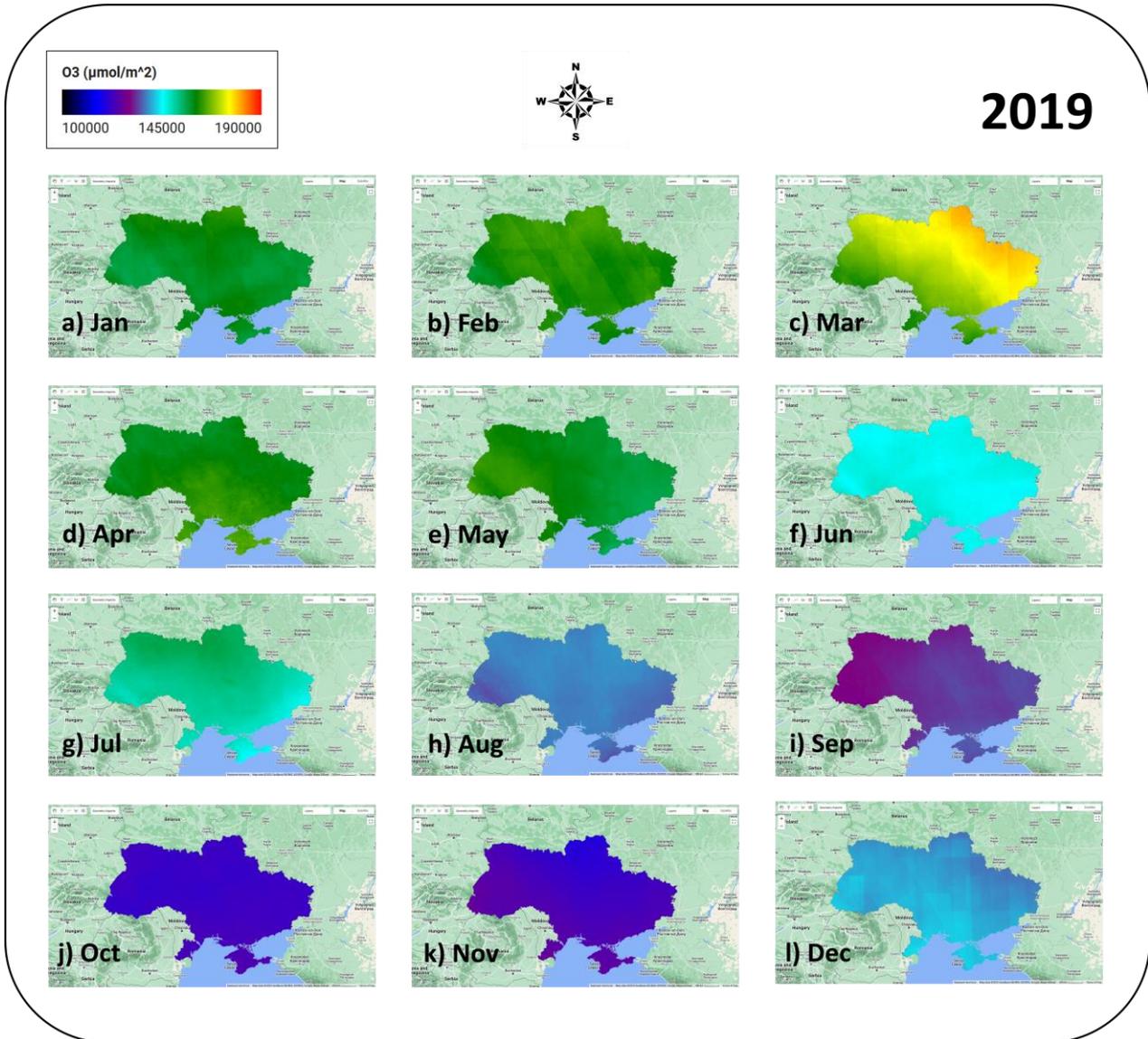
3.2.1 Pollution Observations

The results of this section comprise three items: (i) the concentration maps of the intended pollutant are depicted for 2022 and BAU years, (ii) the annual trends are shown in the form of diagrams (line charts and box plots), and (iii) a correlation assessment using Pearson correlation coefficient (R_P) to address the compatibility between the trends of the three years.

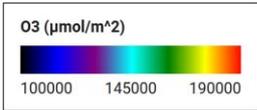
3.2.1.1 O₃

Figure 6 shows the results related to the O₃ concentration. At a glance, it can be seen that there are significant differences between the O₃ concentration, especially for the first six months of each year. Concerning the pre-war time, 2021 began with a much clearer air, compared to 2019 and 2022. As for February, central and western Ukraine have been more polluted in 2022. Taking March as another instance, the southern and northern parts of the country are illustrated with light green and yellow in 2019, while it is almost vice versa in 2022. As for 2021, the map indicates a

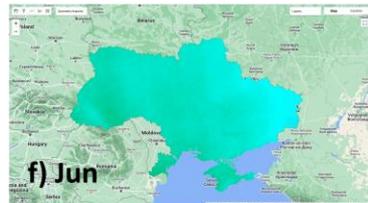
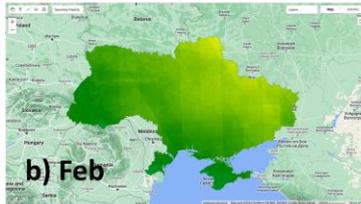
considerably higher concentration of this pollutant all over the country. Similar examples can be Jun and September, for which, the concentration of O₃ in 2022 is tangibly higher than BAU years.



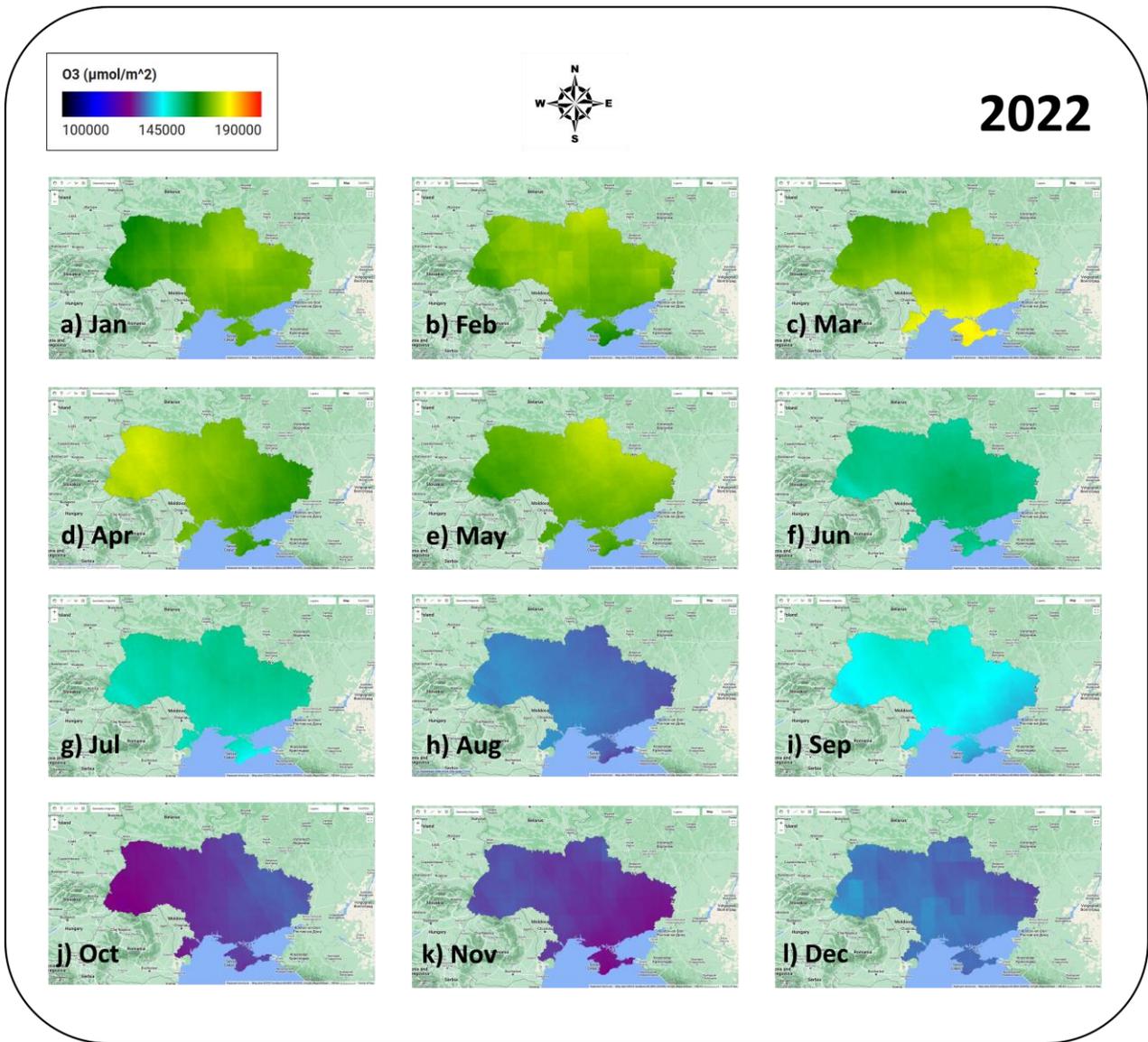
(a)



2021



(b)

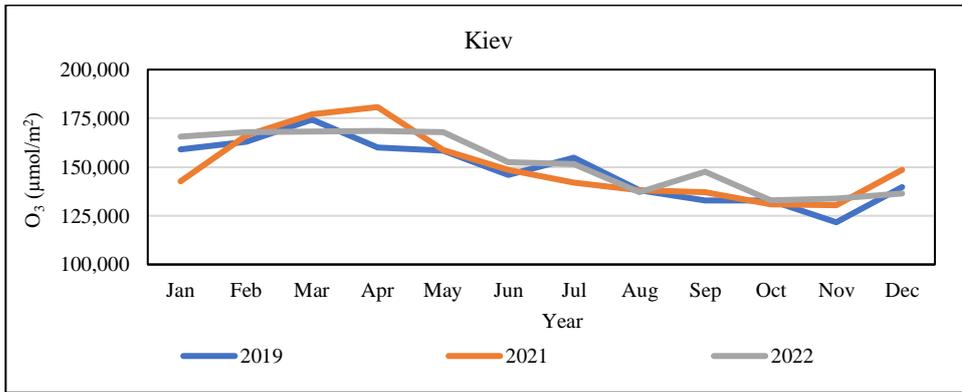


(c)

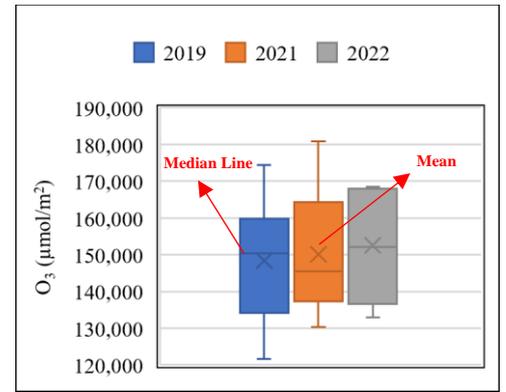
Figure 6: The concentration of O₃ in (a) 2019, (b) 2021, and (c) 2022

Figure 7 depicts annual trends of the O₃ concentration along with the corresponding box plots for five selected sites. Based on these charts, it can be said that for all cities, the O₃ concentration peaks in spring (mostly March), and thereafter it follows a downward trend until November and rises again at the end of the year. This remark is consistent with the maps in Figure 6.

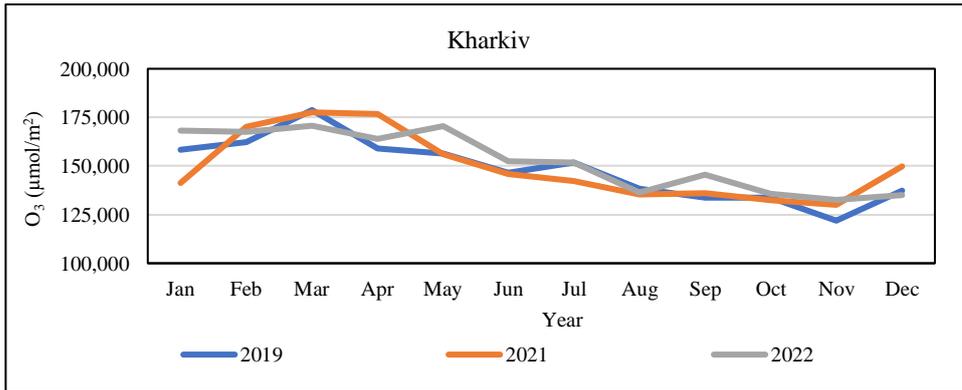
Although the general trend is the same for 2022 and BAU years, there are some minor distinctions. For instance, at the beginning of each chart, the line of 2021 climbs more steeply than the two others. Also, in the last month, the line 2022 represents a gentler inclination. Another appreciable deduction is that in most cases from May to September, the concentrations of O₃ pertaining to 2022 have been above BAU years.



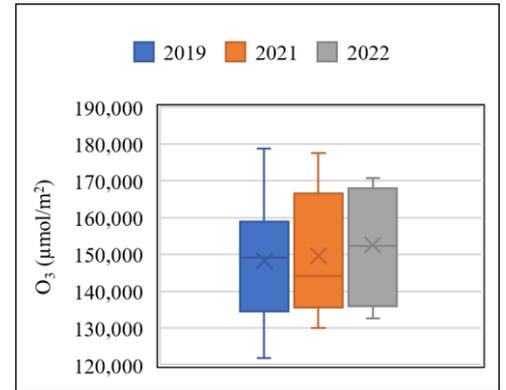
(a)



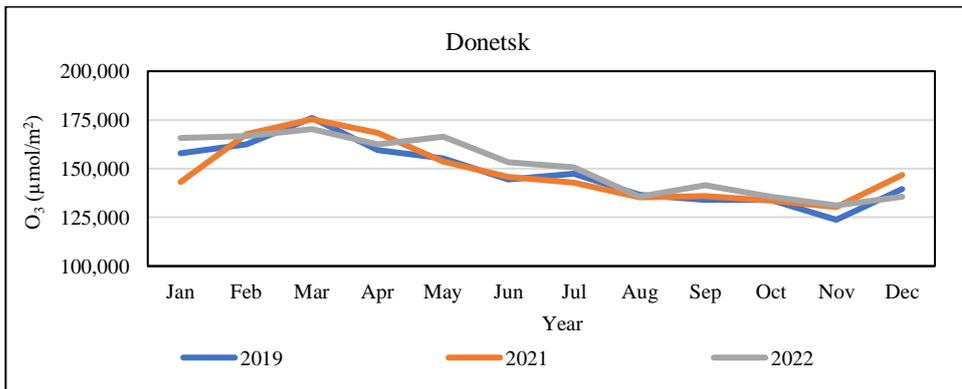
(b)



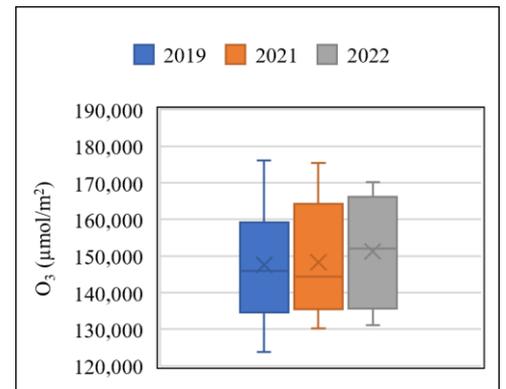
(c)



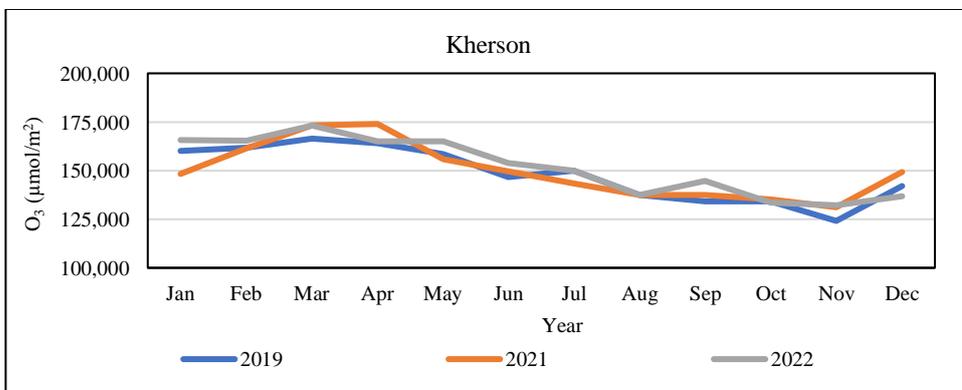
(d)



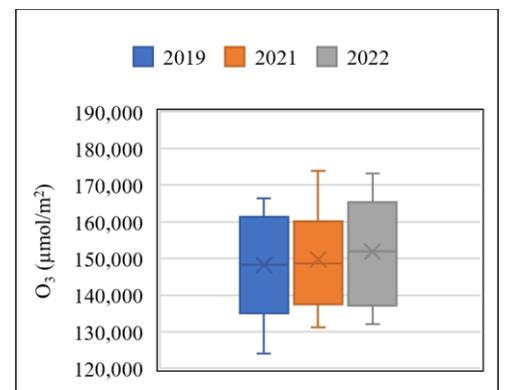
(e)



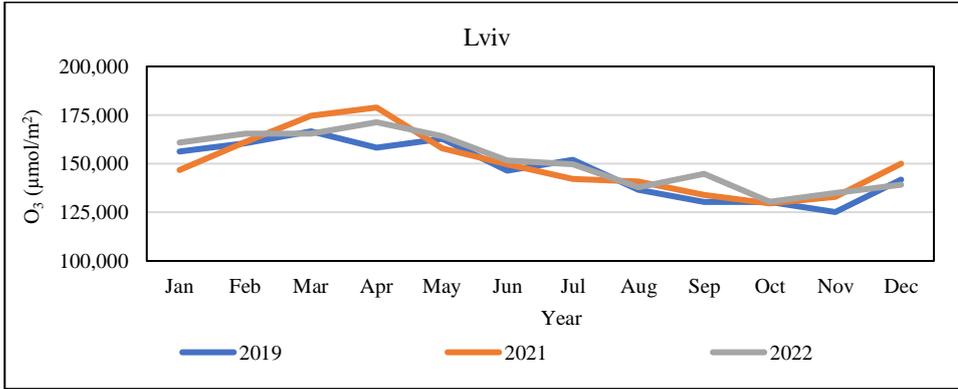
(f)



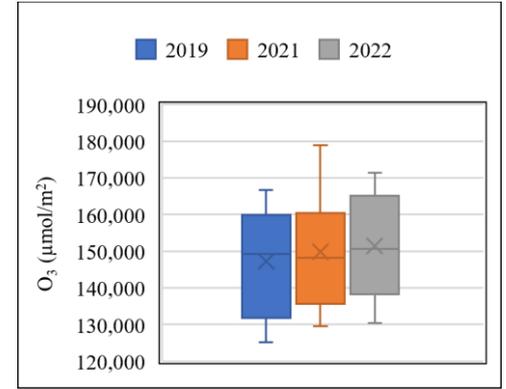
(g)



(h)



(i)



(j)

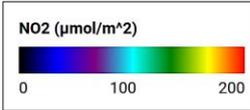
Figure 7: Annual trends and box plots of the O₃ concentration in (a and b) Kiev, (c and d) Kharkiv, (e and f) Donetsk, (g and h) Kherson, and (i and j) Lviv

Table 3 expresses the results of the correlation analysis between the considered years. High R_p values for the five cities across Ukraine indicate the compatibility between the behavior of all three years, as previously inferred from Figure 6 and Figure 7.

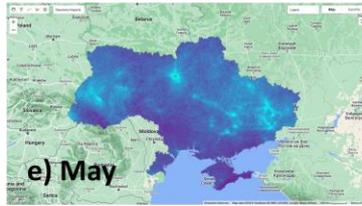
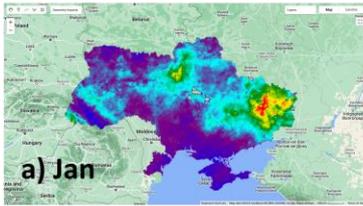
Place	2019 & 2022	2021 & 2022	2019 & 2021
Kiev	0.91	0.80	0.83
Kharkiv	0.91	0.74	0.84
Donetsk	0.94	0.82	0.91
Kherson	0.95	0.85	0.90
Lviv	0.92	0.87	0.85

3.2.1.2 NO₂

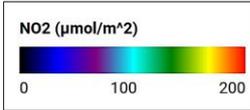
Figure 8 illustrates the maps of the NO₂ concentration. Unlike the O₃ behavior, the differences in January, February, and December maps are the most prominent. In January 2019 and 2021, a large amount of NO₂ was observed in the central parts and capital (i.e., Kiev), while it has been considerably reduced in 2022. A similar observation exists for February 2021 versus 2019 and 2022. Also, the map of November 2022 addresses some increments of the NO₂ emission, especially in western Ukraine. As for December, 2019 had a much clearer sky, particularly in western parts.



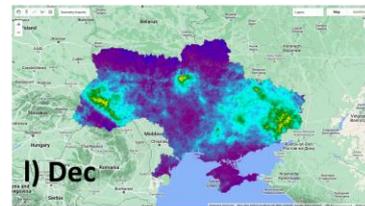
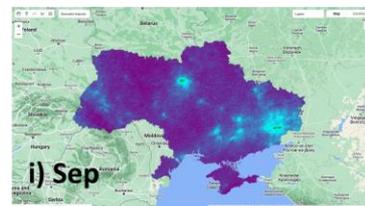
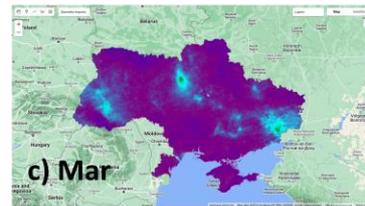
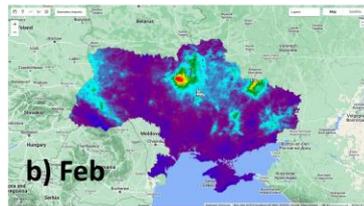
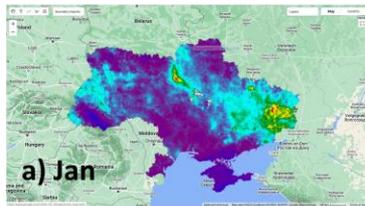
2019



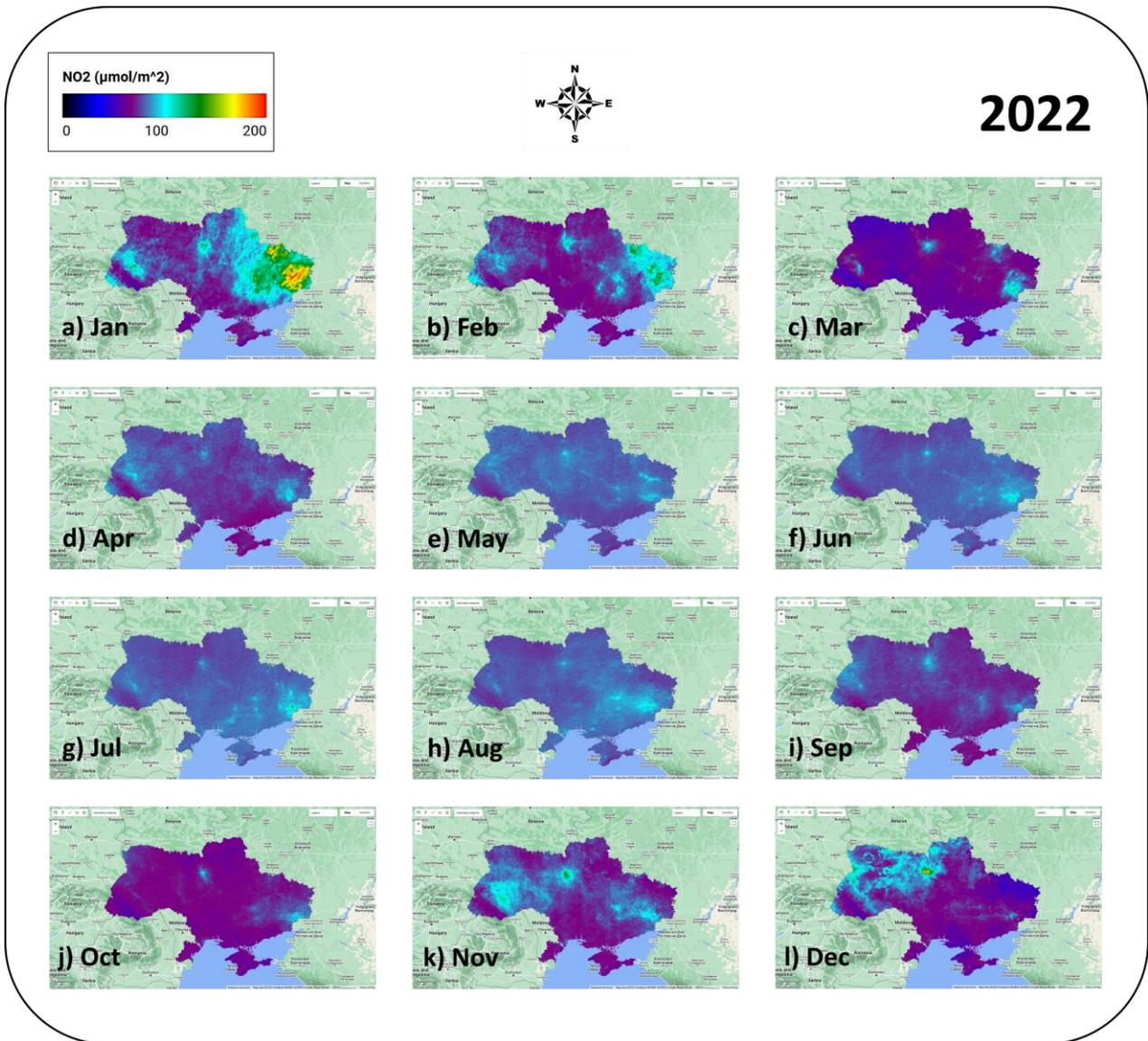
(a)



2021



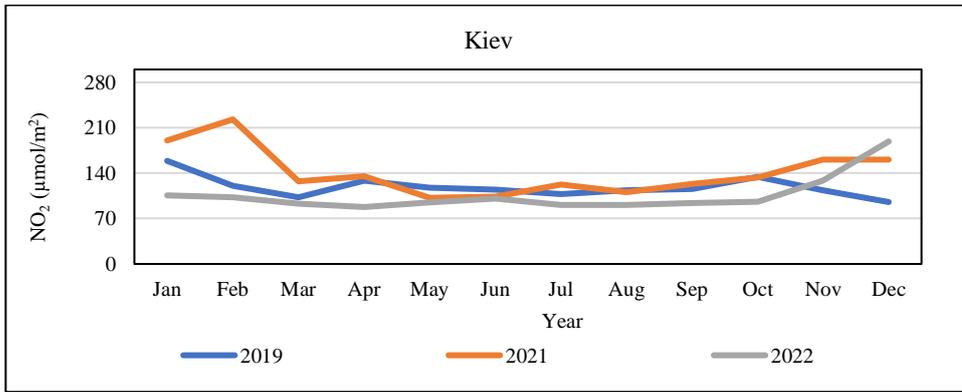
(b)



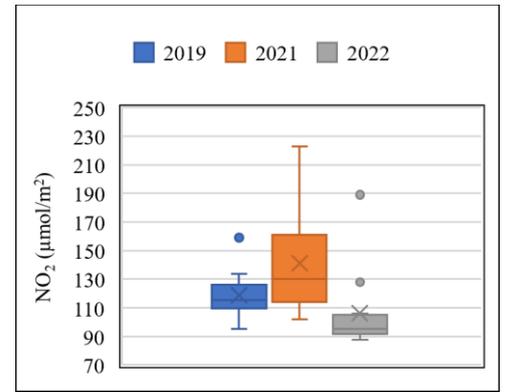
(c)

Figure 8: The concentration of NO₂ in (a) 2019, (b) 2021, and (c) 2022

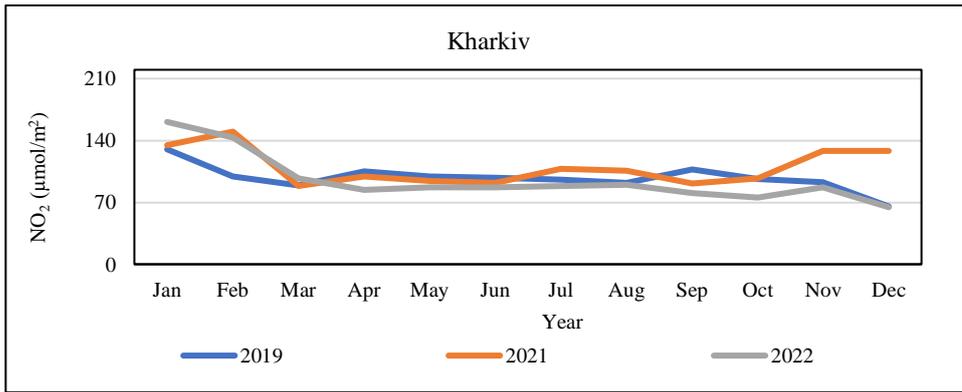
Referring to Figure 9 which compares the annual trends of the NO₂ concentration in five selected sites, there is an ample difference between the middle and extreme parts of all charts. From March to October, steadier behavior is evident compared to the beginning and final months. This inference can be obtained also from the box plots wherein the extreme values are shown by outlier points and it is in agreement with the explanations of Figure 8. For these less turbulent times in Kiev and Kharkiv, the NO₂ concentration in 2022 is lower than BAU time, while the lines are very close and twisted for three other cities.



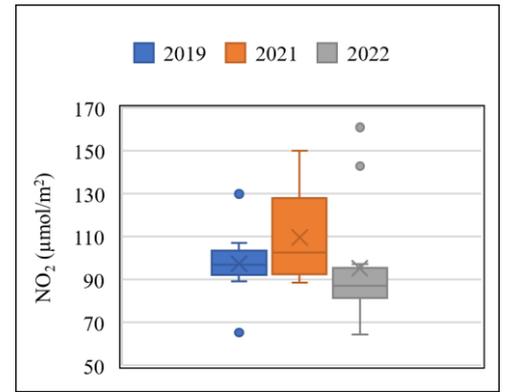
(a)



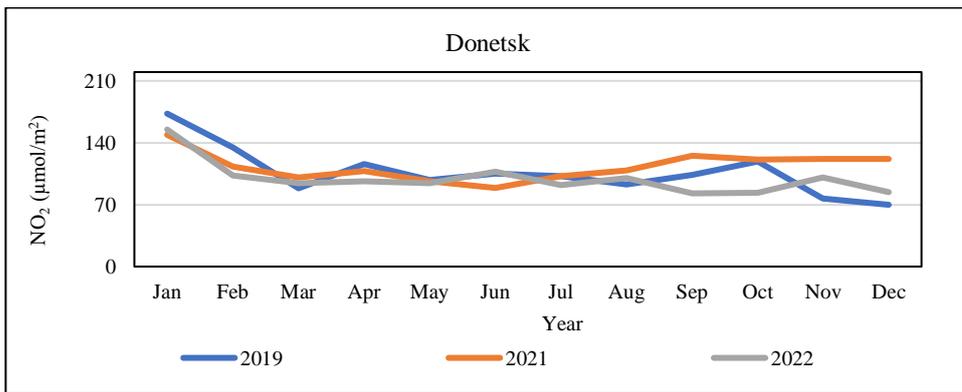
(b)



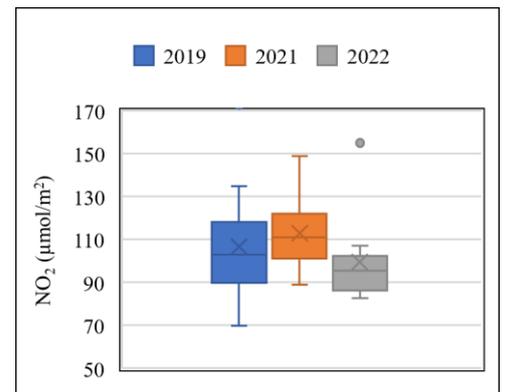
(c)



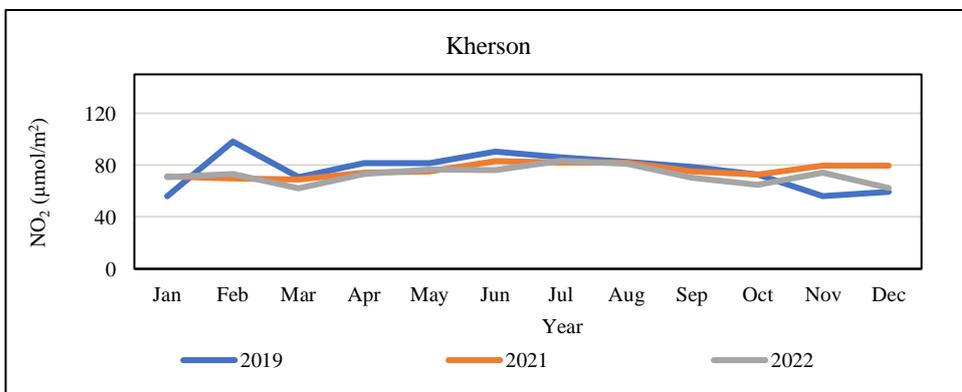
(d)



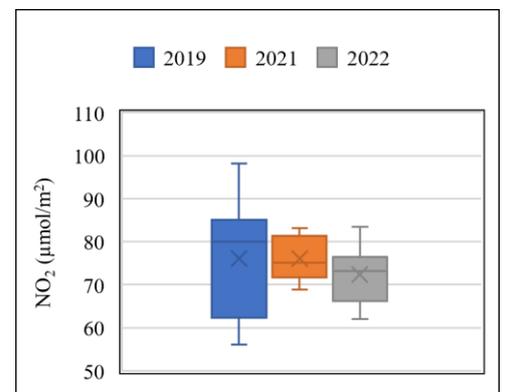
(e)



(f)



(g)



(h)

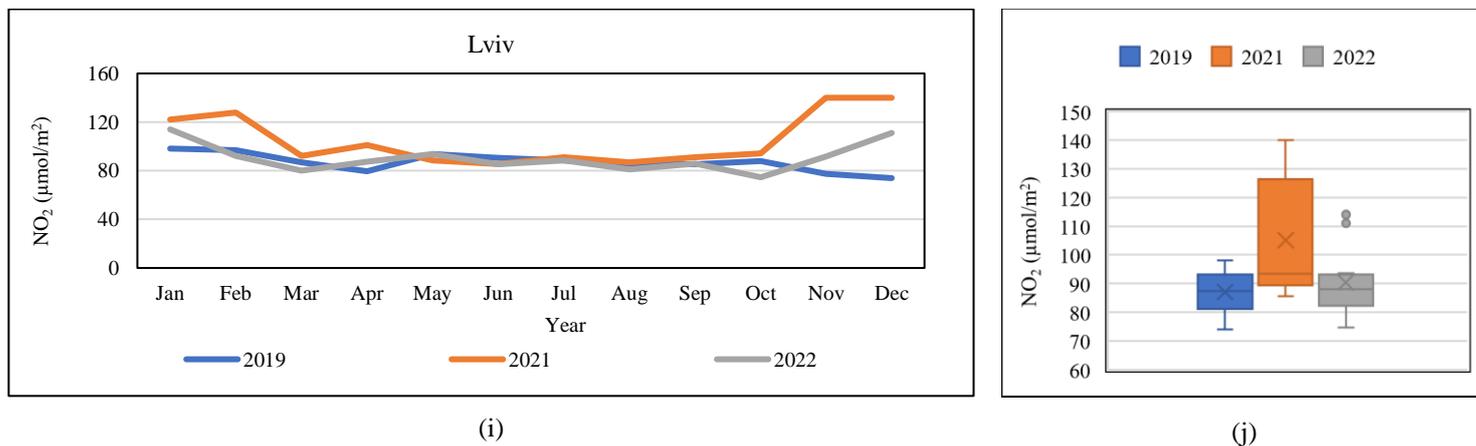


Figure 9: Annual trends and box plots of the NO₂ concentration in (a and b) Kiev, (c and d) Kharkiv, (e and f) Donetsk, (g and h) Kherson, and (i and j) Lviv

Based on Table 4 which presents the correlation results, there are very low values of R_P and also two negative values that demonstrate meaningful variations of the NO₂ concentration when moving from 2019 and 2021 to 2022. For instance, R_P equals 0.03 for Lviv when comparing 2019 and 2022 records, implying the effect of war on the air quality of this area in terms of NO₂ concentration. On the contrary for Kharkiv, Donetsk, and Kherson, the R_P values in the two first columns are of a higher magnitude which means the annual trend of the NO₂ concentration in 2022 was not much different from BAU years.

Table 4: Yearly correlation results (R_P) of NO₂ for the considered cities

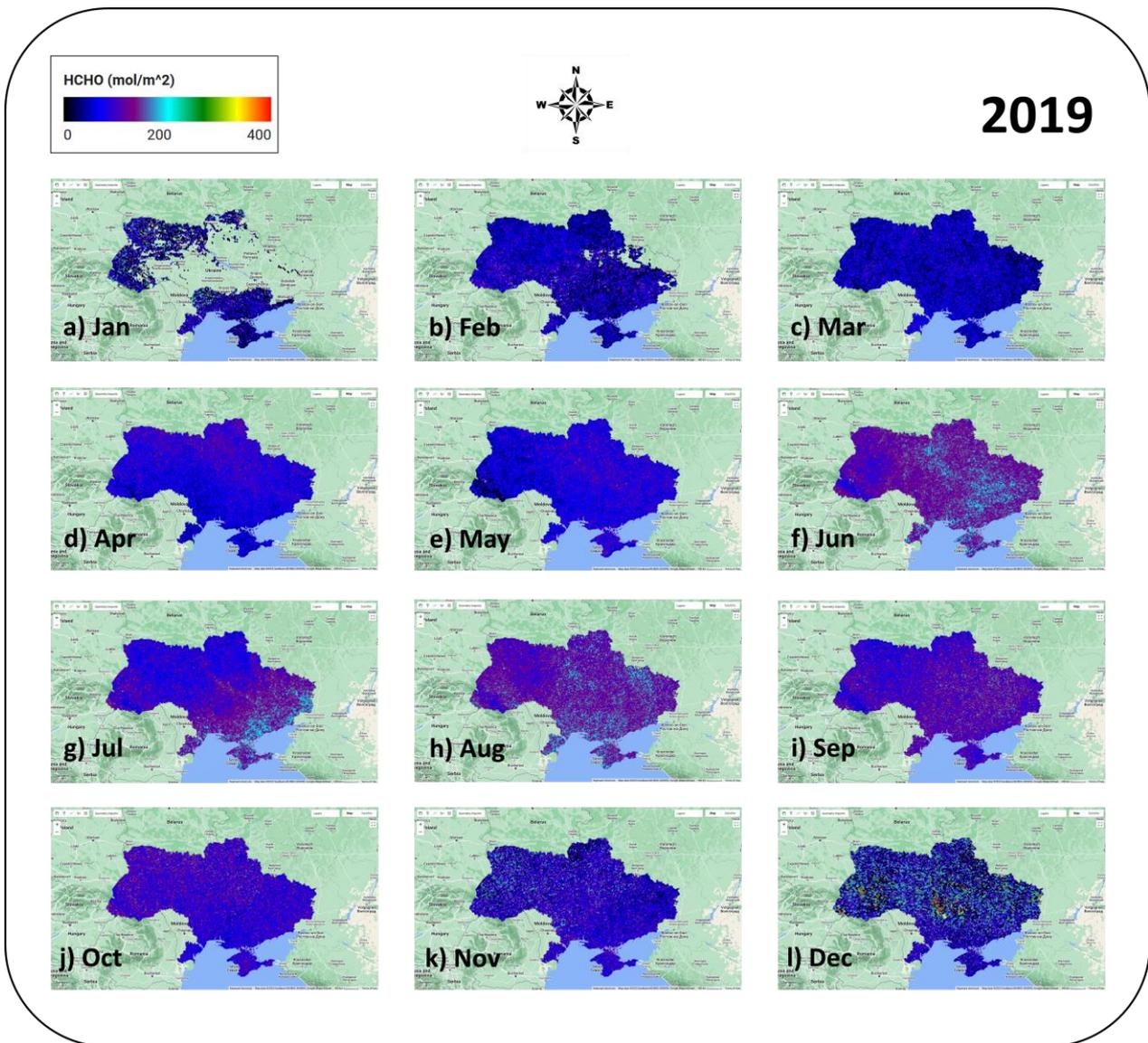
Place	2019 & 2022	2021 & 2022	2019 & 2021
Kiev	-0.37	0.33	0.34
Kharkiv	0.68	0.60	0.05
Donetsk	0.74	0.48	0.48
Kherson	0.50	0.57	0.07
Lviv	0.03	0.67	-0.29

3.2.1.3 HCHO

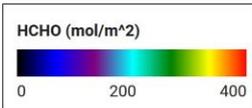
It was explained that observations acquired for the CH₄ and SO₂ contained a large number of negative values. A few negative values also existed in the HCHO records, which were eliminated for all statistical analyses in the first place. Likewise, there were missing records for this pollutant, especially for the initial and final months of all years. These issues explain the inconsistencies in the maps and charts related to the HCHO.

Figure 10 shows the monthly concentrations of HCHO. According to this figure, and by bounding the assessment to months with complete records, the most remarkable distinction can be found from June to September. Having August as an example, southern and eastern parts of Ukraine have

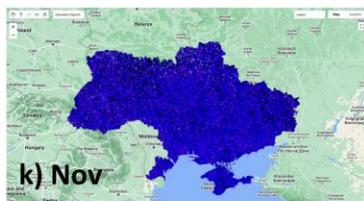
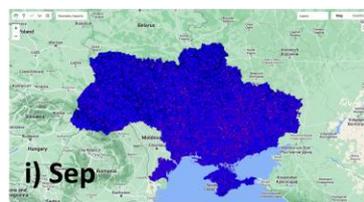
witnessed higher pollution in 2022 compared to previous years. Similarly, the HCHO emission in western Ukraine in July has increased in 2022.



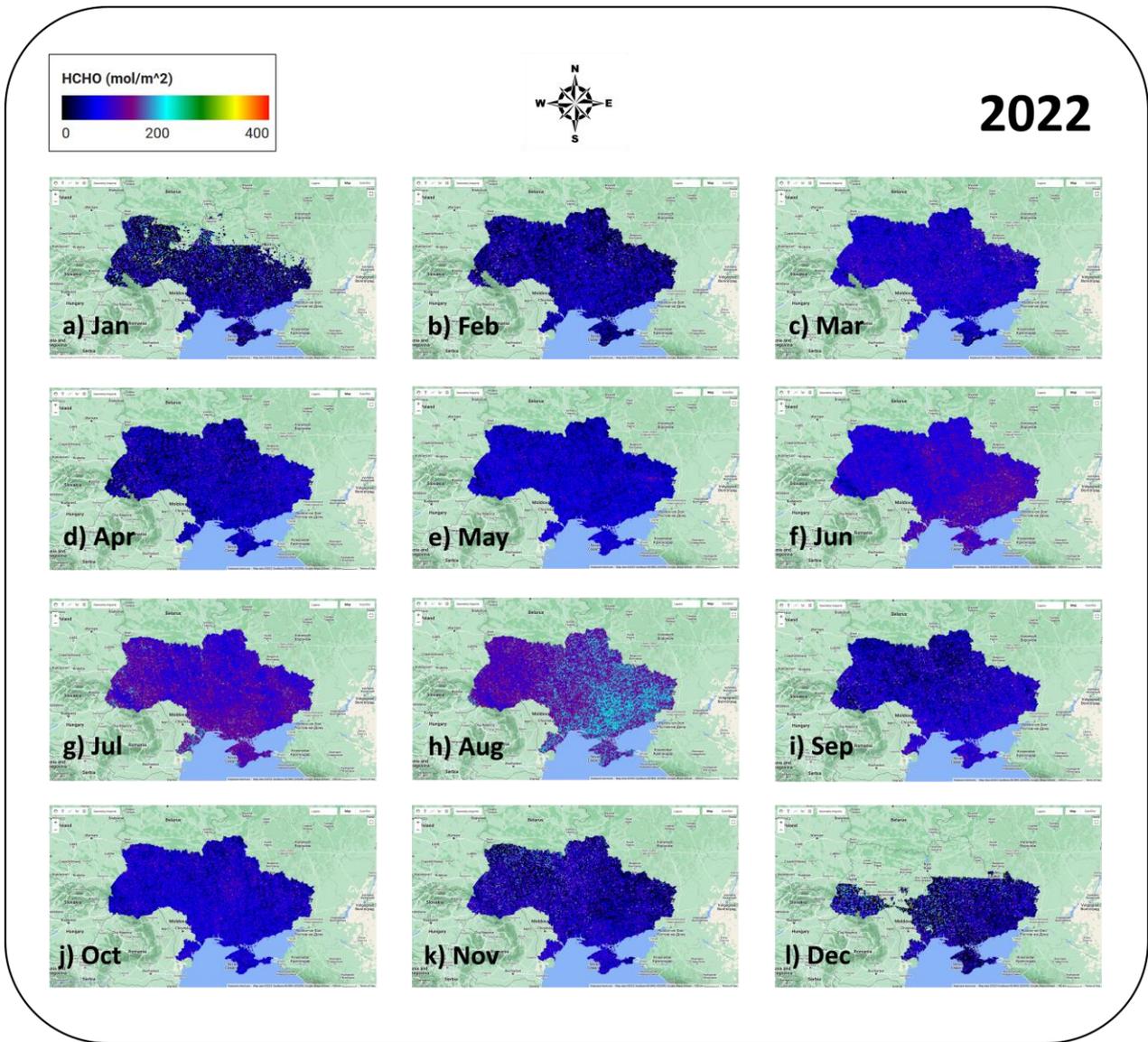
(a)



2021



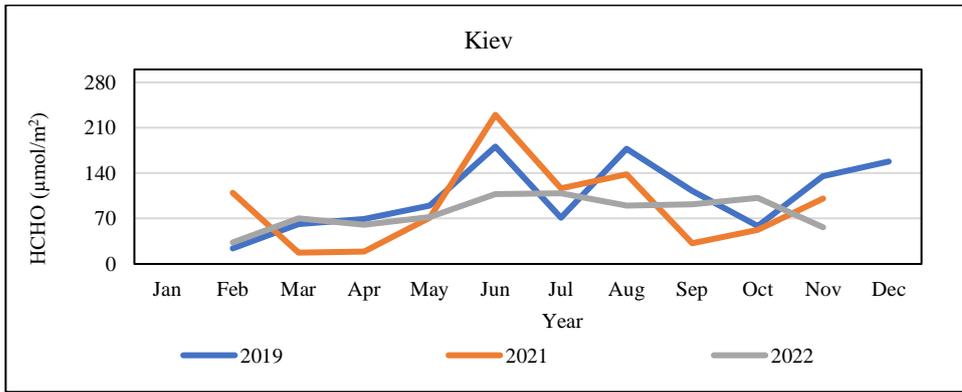
(b)



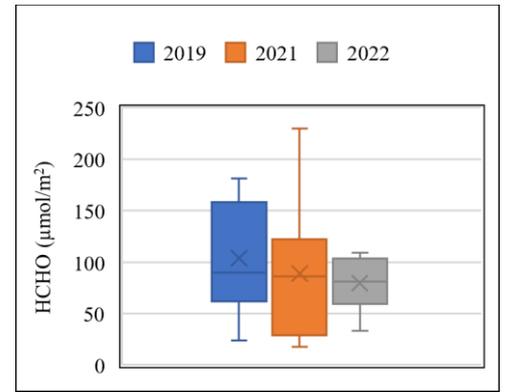
(c)

Figure 10: The concentration of HCHO in (a) 2019, (b) 2021, and (c) 2022

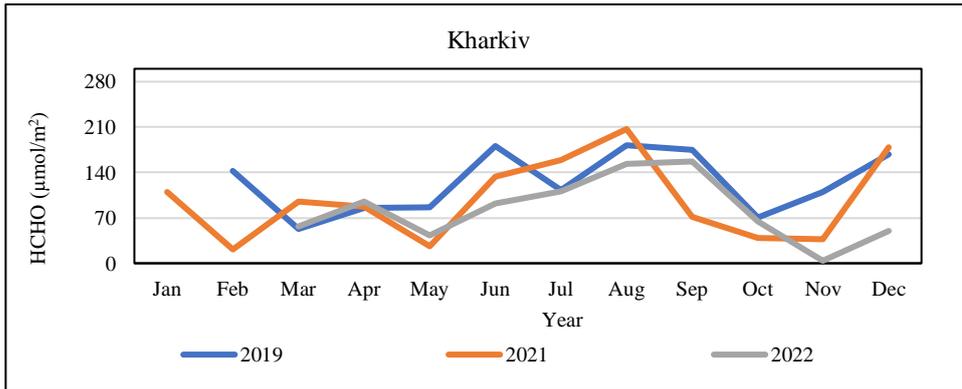
Figure 11 shows and compares the annual trends of the HCHO concentration in five selected sites besides the box plots. As is seen, there are lots of fluctuations in the left charts, and also, due to the lack of few records, it is hard to deduce a general meaningful trend from them. However, in accordance with the above explanations, it is possible to identify some peak points in August and September for Kharkiv, Donetsk, and Kherson.



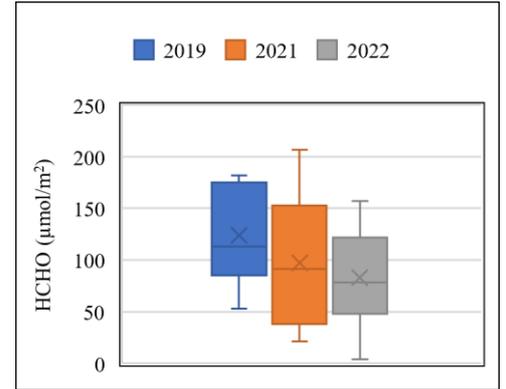
(a)



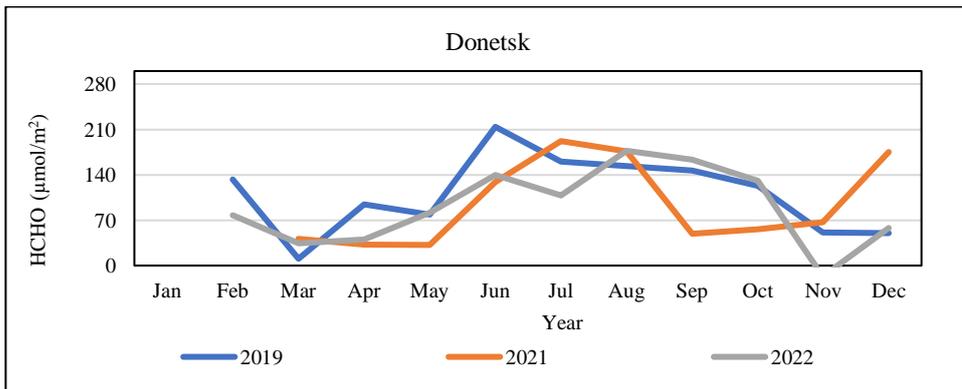
(b)



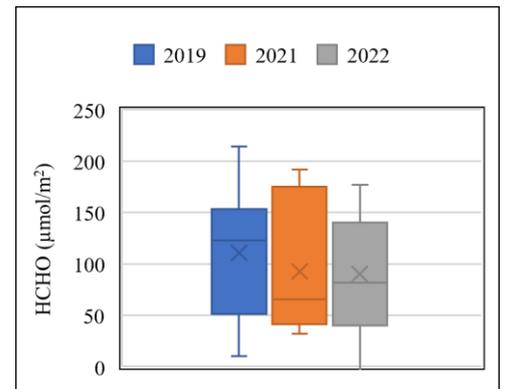
(c)



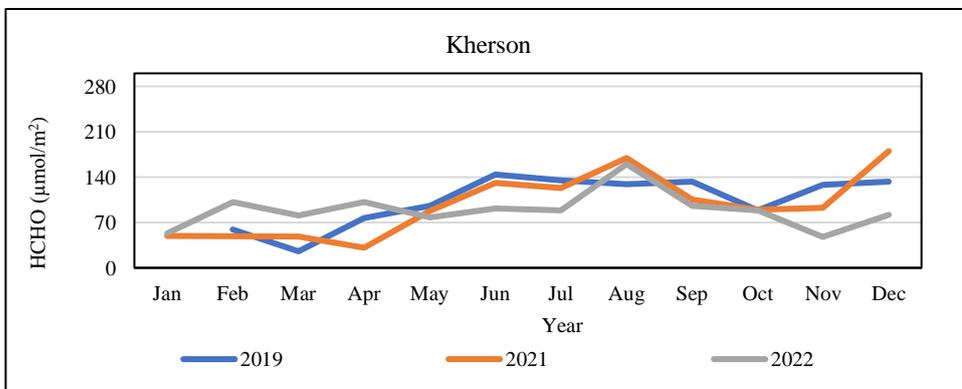
(d)



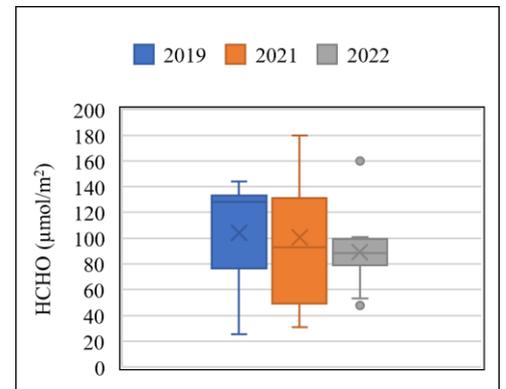
(e)



(f)



(g)



(h)

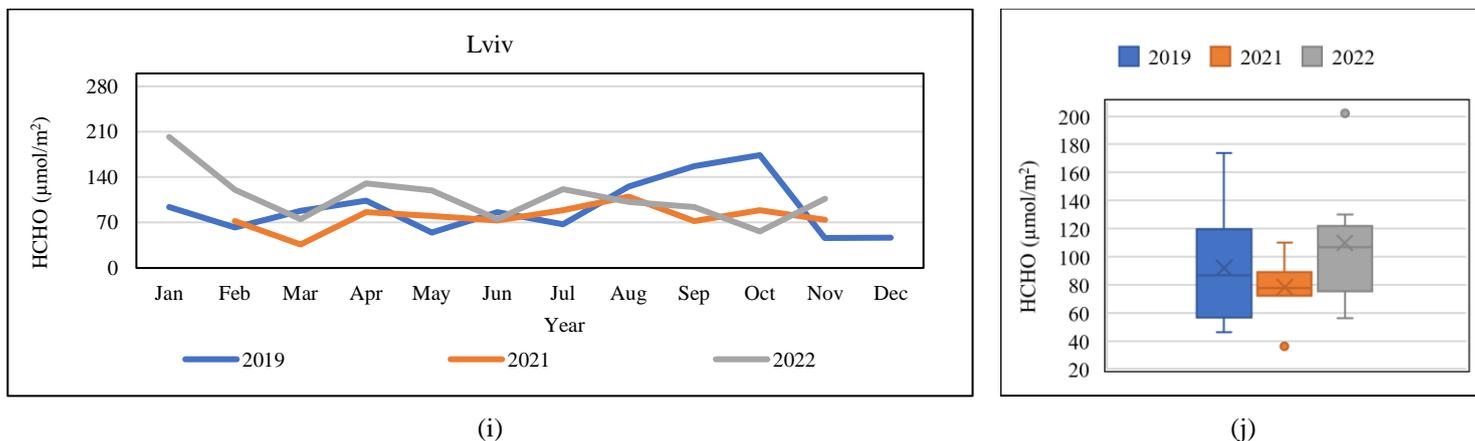


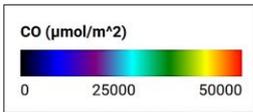
Figure 11: Annual trends and box plots of the HCHO concentration in (a and b) Kiev, (c and d) Kharkiv, (e and f) Donetsk, (g and h) Kherson, and (i and j) Lviv

As is denoted in Table 5, R_P values are mostly ununiform for each city and each pair of years. However, the results support previous deductions such as the adverse correlation of 2019 and 2022 trends in Lviv. From a more general point of view, concerning the comparison of 2022 with BAU years, the variation of R_P in the second column is less than in the first column. Given that 2021 was affected by the COVID-19 lockdown, it can be said that evident changes occurred in the air quality in terms of HCHO when 2022 is evaluated versus 2019.

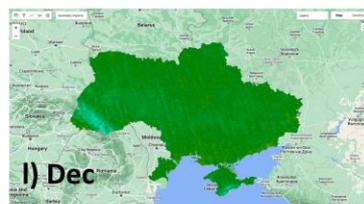
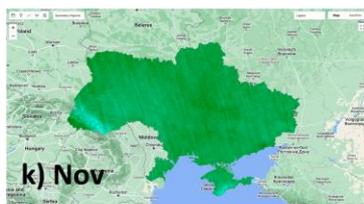
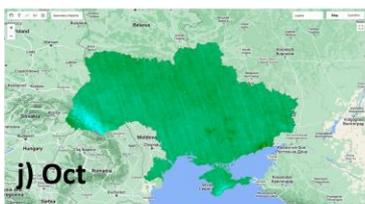
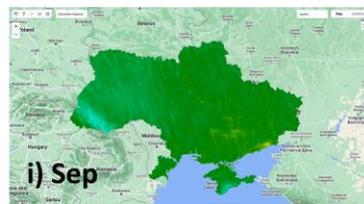
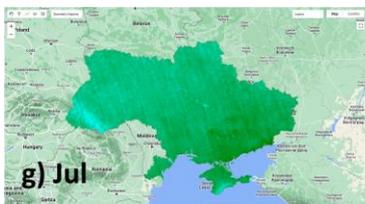
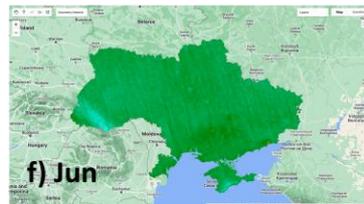
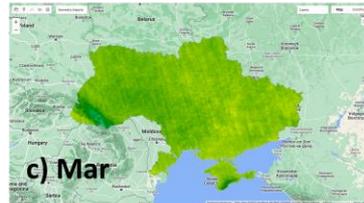
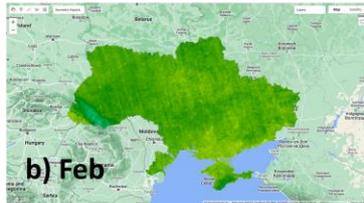
Place	2019 & 2022	2021 & 2022	2019 & 2021
Kiev	0.44	0.32	0.62
Kharkiv	0.53	0.49	0.61
Donetsk	0.79	0.34	0.40
Kherson	0.12	0.36	0.75
Lviv	-0.58	0.28	0.25

3.2.1.4 CO

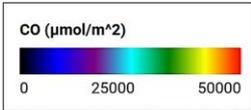
Similar to the O_3 and NO_2 , the entire CO records are functional for this research. The maps of the CO concentration are exhibited in Figure 12. Major differences can be detected for the first four months of each year, as well as August, and to some extent September. In contrast, the CO-based pollution of the last three months is roughly similar.



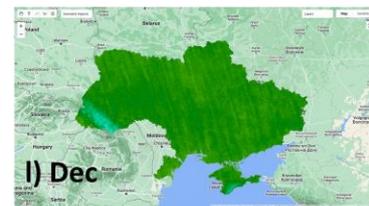
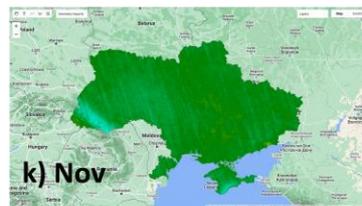
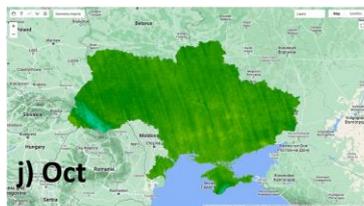
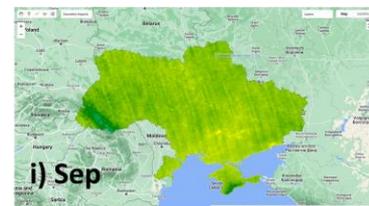
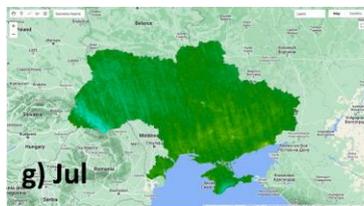
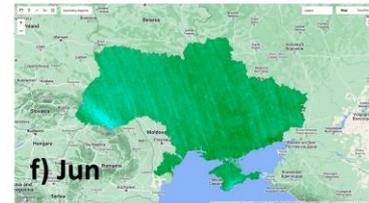
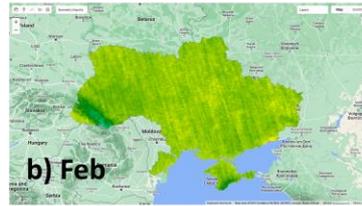
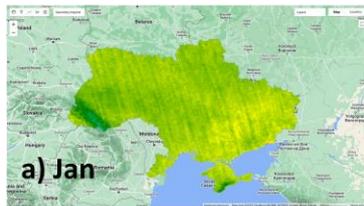
2019



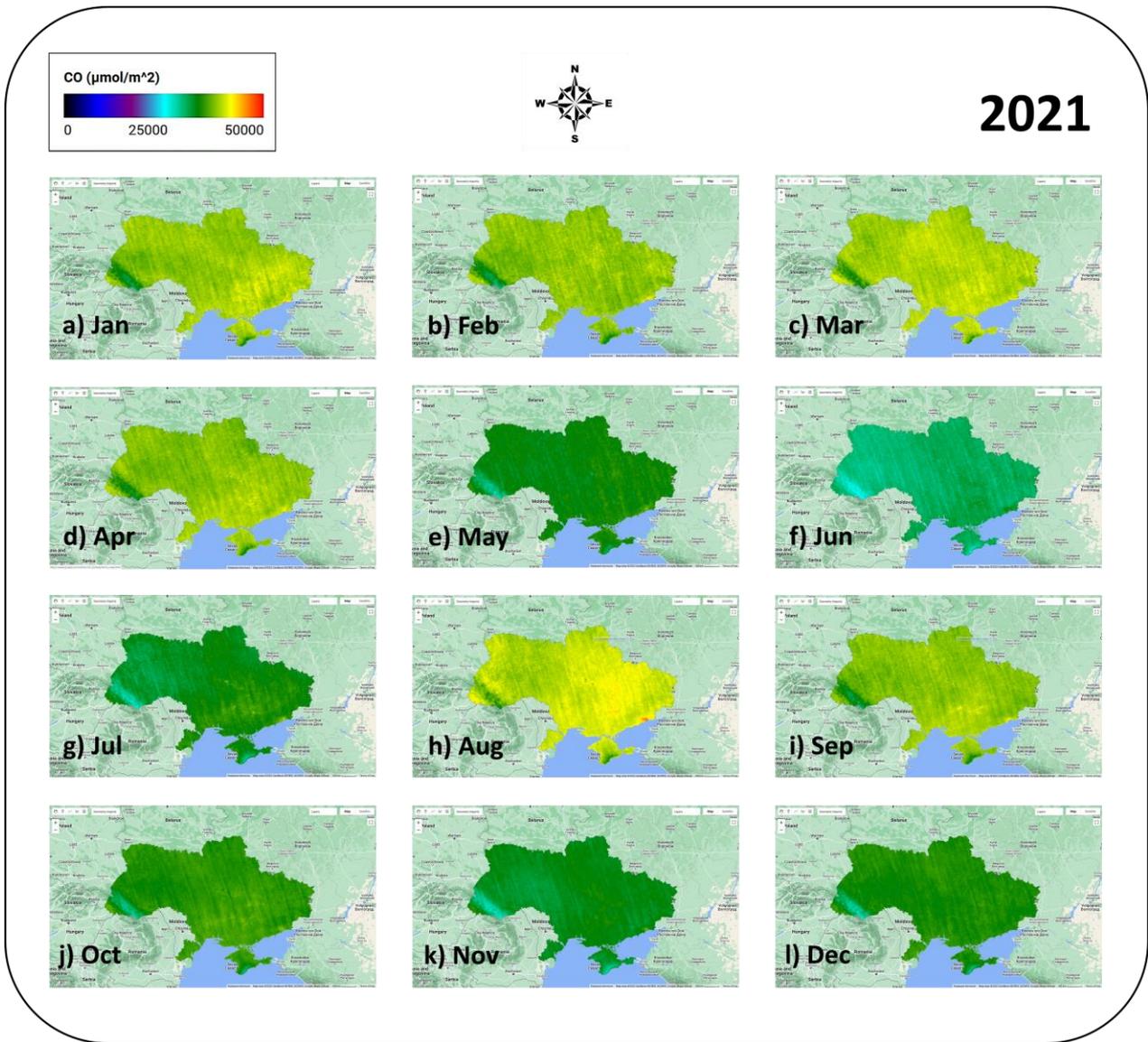
(a)



2021



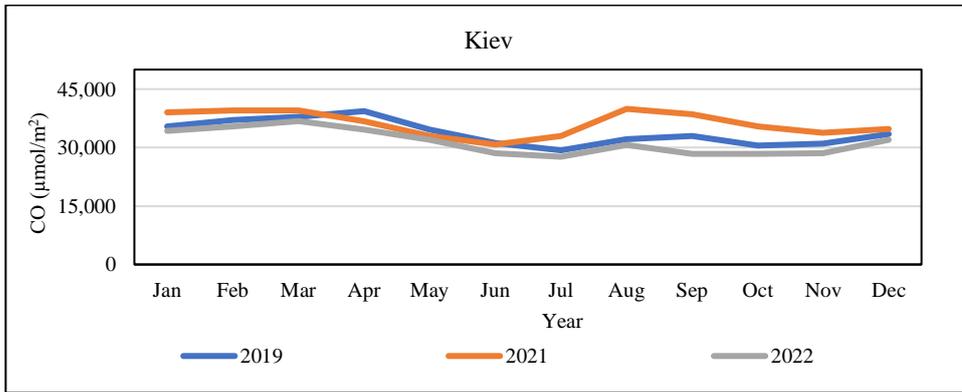
(b)



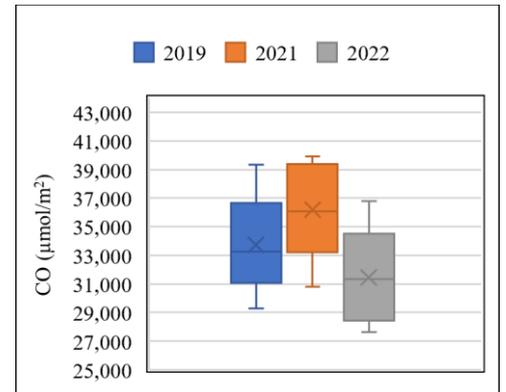
(c)

Figure 12: The concentration of CO in (a) 2019, (b) 2021, and (c) 2022

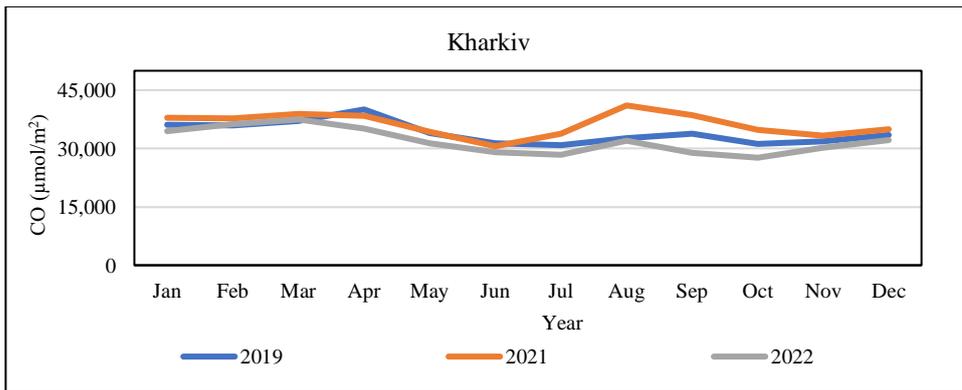
The annual trends of the CO concentration in five selected cities and their box plots are shown in Figure 13. Since the records of 2022 are below BAU years almost in all cases, it can be derived that the war has reduced the CO emission level throughout Ukraine. These charts also exhibit a considerably higher CO concentration in August 2021.



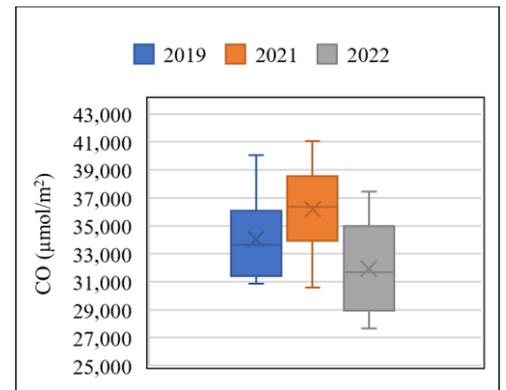
(a)



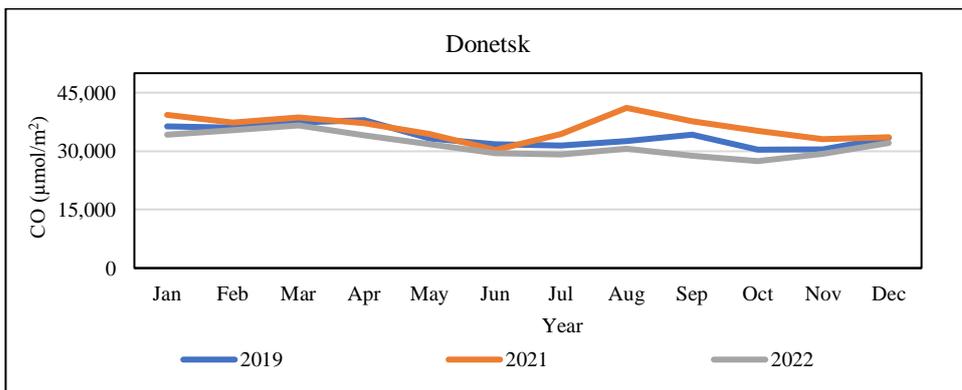
(b)



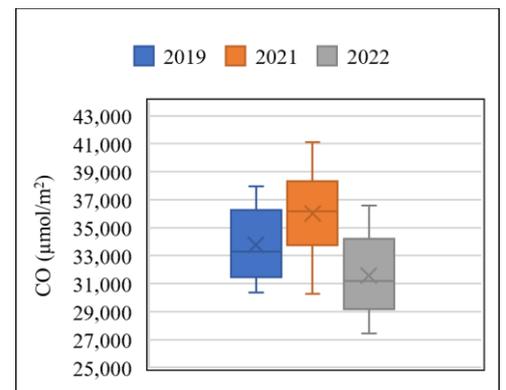
(c)



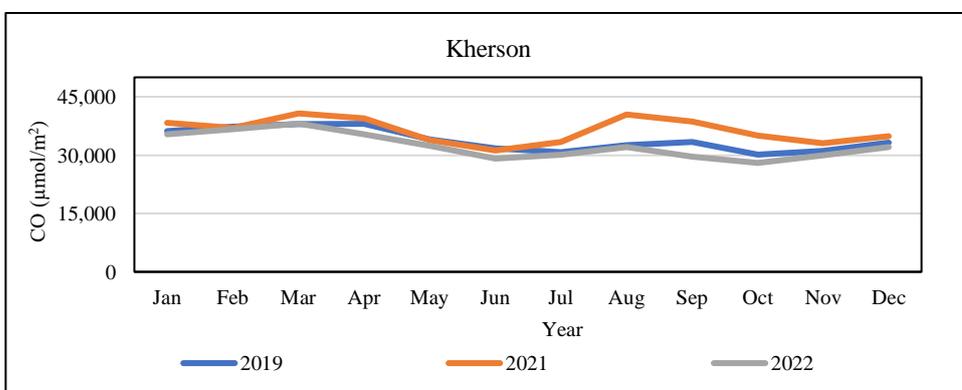
(d)



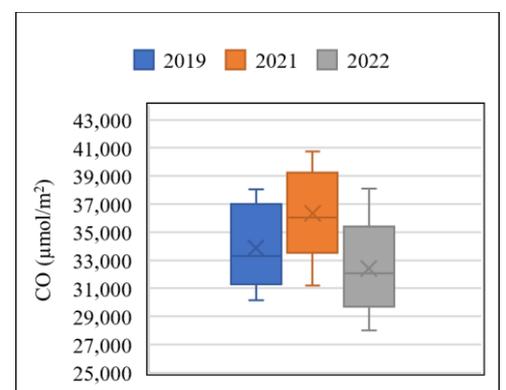
(e)



(f)



(g)



(h)

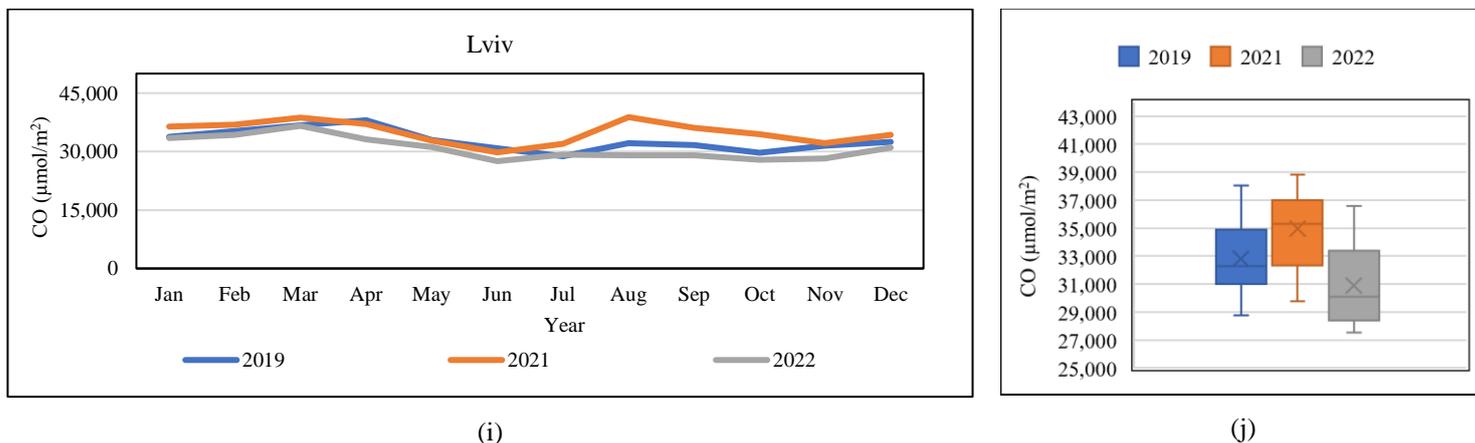


Figure 13: Annual trends and box plots of the CO concentration in (a and b) Kiev, (c and d) Kharkiv, (e and f) Donetsk, (g and h) Kherson, and (i and j) Lviv

Referring to the R_P values given in Table 6, there is a large harmony between the CO emission in 2022 and 2019 for the whole Ukraine. Moreover, lower magnitudes of R_P for 2021 against both 2022 and 2019 indicate a more different emission pattern for this year.

Table 6: Yearly correlation results (R_P) of CO for the considered cities

Place	2019 & 2022	2021 & 2022	2019 & 2021
Kiev	0.92	0.59	0.55
Kharkiv	0.85	0.58	0.58
Donetsk	0.89	0.45	0.57
Kherson	0.94	0.63	0.65
Lviv	0.85	0.63	0.63

3.2.2 Air Quality at the Beginning of the War

This section points out the changes in air quality during the initial days of the Russia-Ukraine crisis. Needless to say, declaring an exceptional situation such as war affects many usual patterns of anthropogenic activities including human circulation and industrial processes. Hence, the emission of O_3 , NO_2 , HCHO, and CO is monitored from late February-early March. The data of this section are obtained from Earth Engine Apps TROPOMI Explorer [88] as it provides air quality records in the form of line charts which are suitable for comparison purposes. Figure 14 shows a view of this portal.

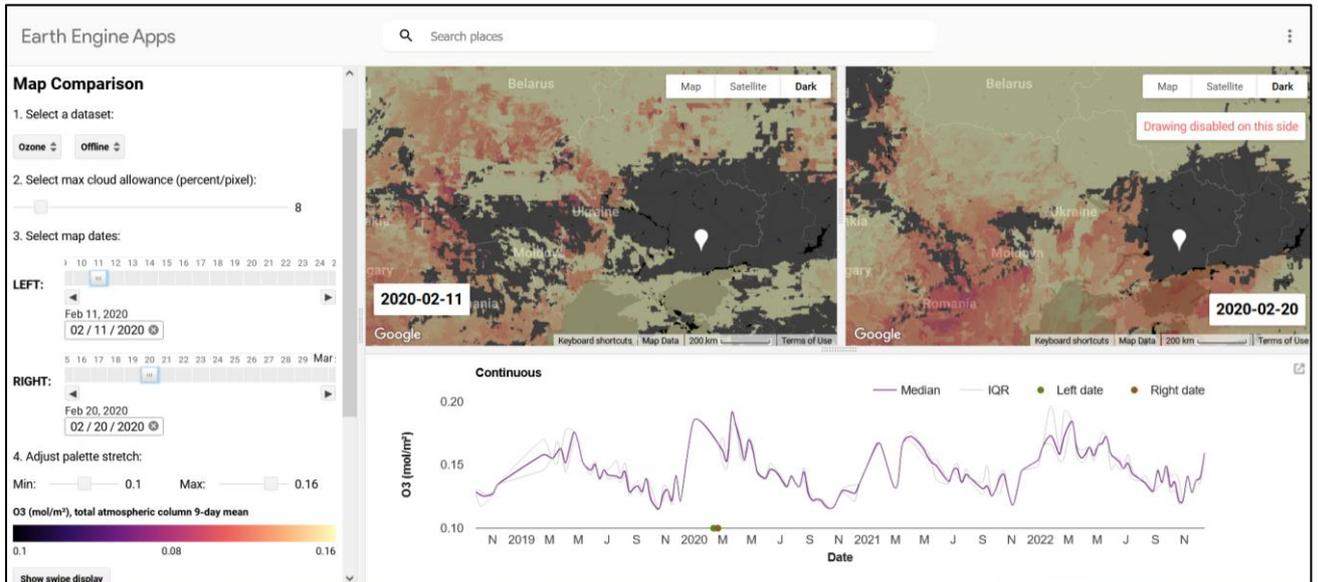
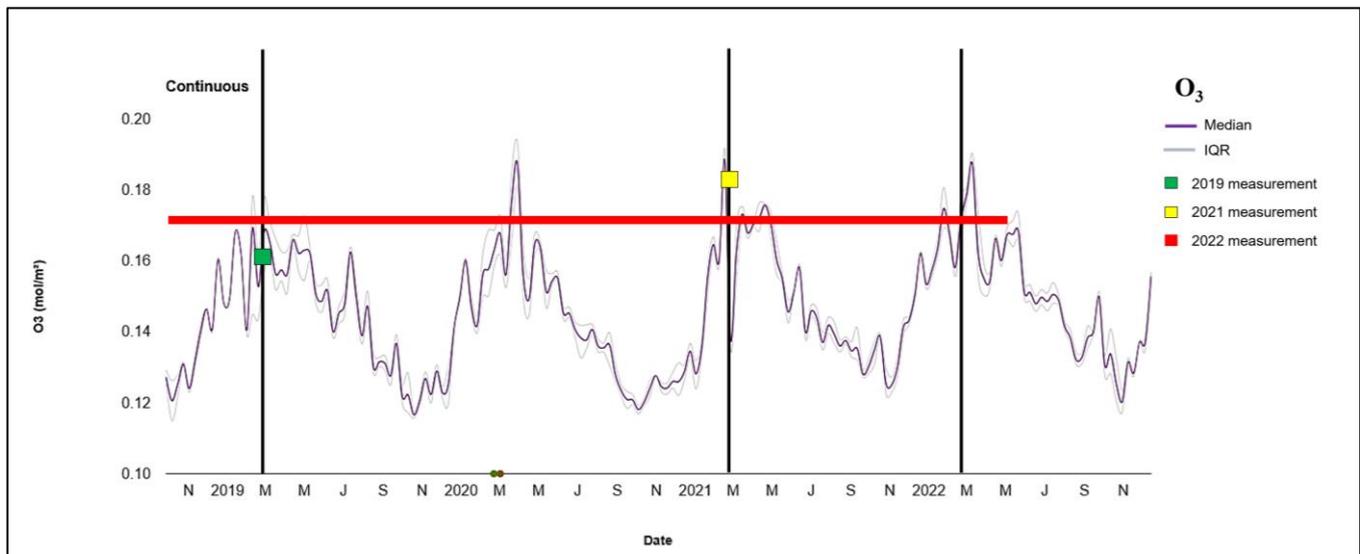
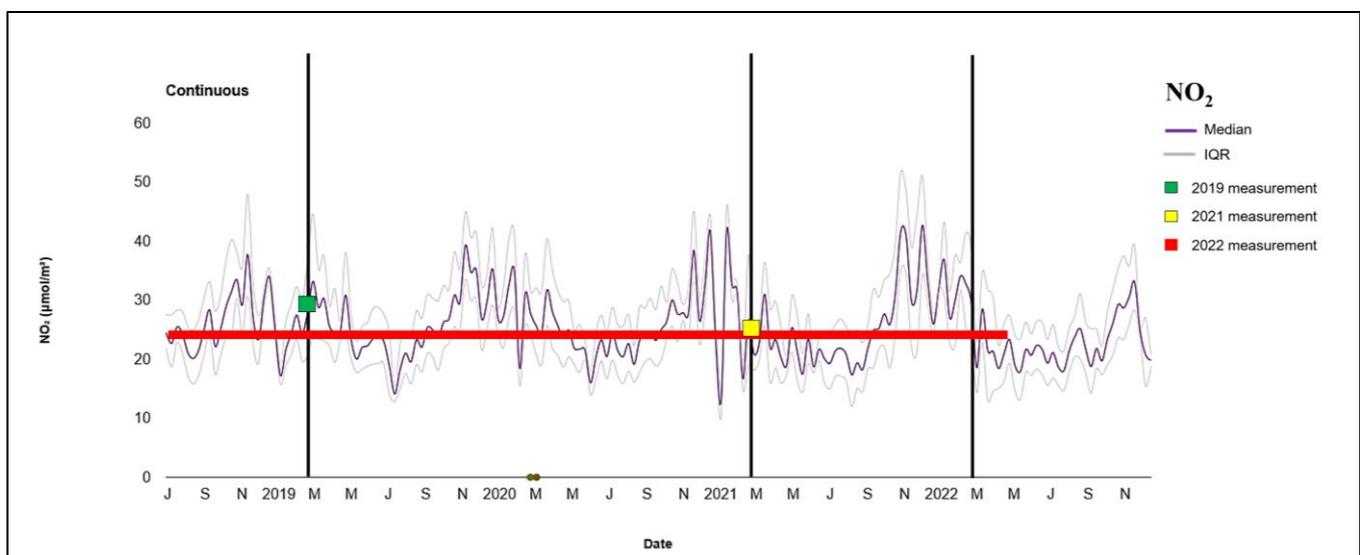


Figure 14: A view of the TROPOMI Explorer webpage [114]

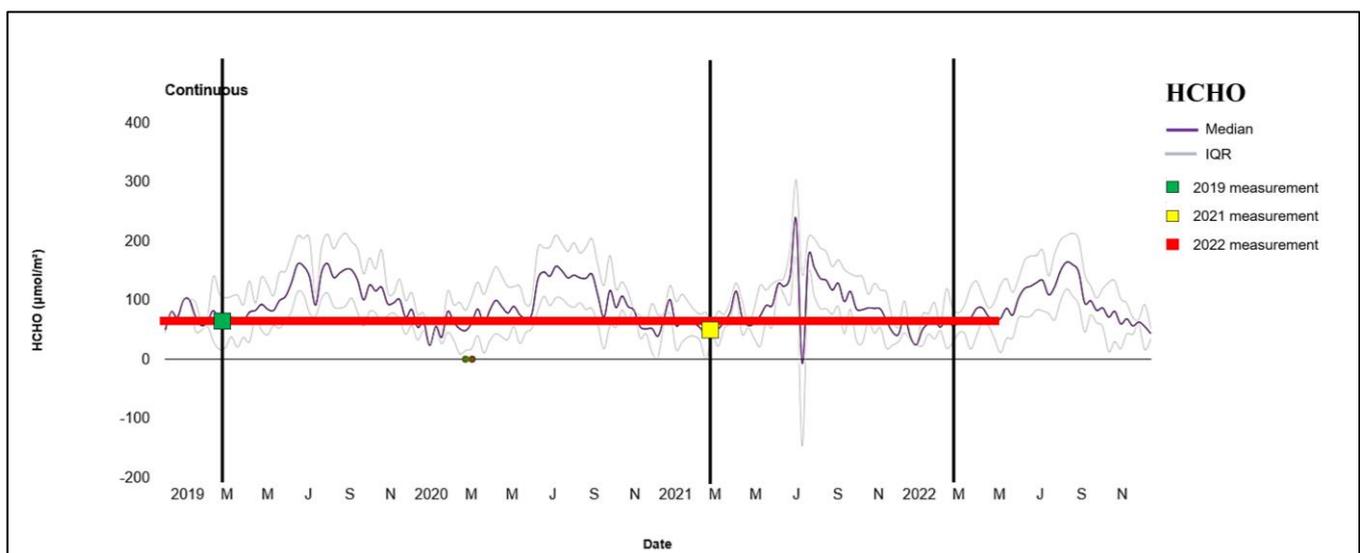
Imposing Ukraine as the AOI, a chart is created for each pollutant using the 9-day mean concentrations. Hereupon, the considered period is February 24 - March 05. Figure 15 depicts the obtained charts wherein the intended period for 2019, 2021, and 2022 is elucidated with black vertical lines and the concentration of each pollutant is determined where this line crosses the chart line. Nevertheless, the observation of 2022 is stretched all over the chart to better compare it with BAU measurements.



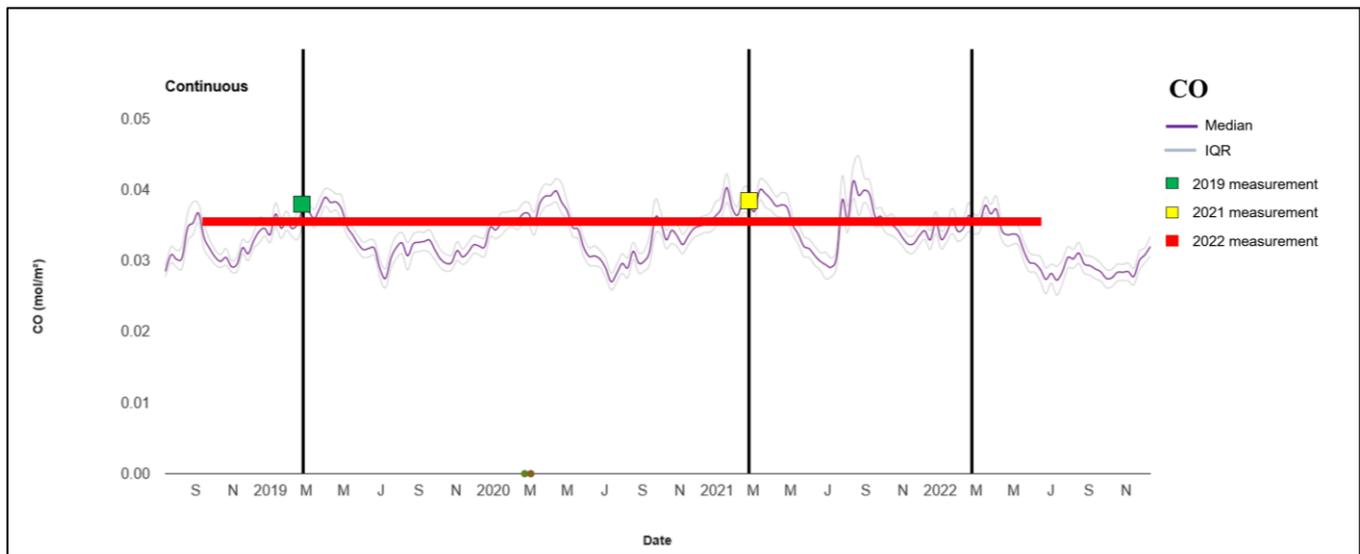
(a) (Ordinate: O₃ total atmospheric column)



(b) (Ordinate: NO₂ tropospheric vertical column)



(c) (Ordinate: HCHO tropospheric column)



(d) (Ordinate: CO vertically integrated column)

Figure 15: Changes in the pollutant emissions (a) O₃, (b) NO₂, (c) HCHO, and (d) CO in the first week of the war

For the case of O₃, NO₂, and CO, the emissions in 2021 are above 2022, while the emission of HCHO is slightly lower. As for 2019, the concentration of NO₂ and CO surpasses 2022, while O₃ has a smaller record and HCHO is more or less of the same concentration. Hence, it can be said that the extreme conditions resulted in lessening the concentration of CO and NO₂ in the initial days of the war. It is also worth noting that the HCHO and O₃ emissions have experienced the lowest and largest changes with respect to BAU years, respectively.

A previous study that investigated the same subject, but for the first two weeks of the war (i.e., February 22 - March 8, 2019-2022) was conducted by [Zalakeviciute, Mejia \[67\]](#). Referring to Table 1 of the cited article, it can be found that their results corroborate what was shown in Figure 15-(b) and (d). However, it is not held for O₃, which could be due to the difference between the examined timespans (Figure 15 is plotted for February 24 - March 05).

3.2.3 Statistical Analysis and Discussion

Focusing on the results presented in Section 3.2.1, for each city, the annual average concentrations of the O₃, NO₂, HCHO, and CO pollutants in 2022 and BAU years are calculated and compared to highlight the changes. The results are given in Table 7. According to this table, each pollutant shows a different behavior based on the location and considered period of assessment.

Table 7: Analysis of pollutant concentrations and changes

Type	Area	O ₃			NO ₂			HCHO			CO		
		2019	2021	2022	2019	2021	2022	2019	2021	2022	2019	2021	2022
Annual Averages (μmol/m ²)	Kiev	148376.59	150057.20	152501.11	118.79	141.17	106.25	103.63	88.91	79.30	33748.67	36171.17	31444.38
	Kharkiv	148169.70	149502.14	152537.40	97.47	109.78	95.43	124.19	97.19	82.67	34025.06	36198.13	31923.14
	Donetsk	147557.95	148231.80	151238.92	106.73	113.10	99.49	110.45	92.31	101.07	33758.81	35990.87	31589.78
	Kherson	148255.17	149691.53	151919.19	76.13	76.13	72.35	104.30	100.58	89.02	33863.69	36333.37	32411.45
	Lviv	147215.15	149849.13	151365.59	87.09	105.11	90.47	92.09	78.32	109.57	32808.89	34956.88	30871.74
Changes of 2022 WRT 2019 and 2021	Kiev	+2.78	+1.63	-	-10.56	-24.73	-	-23.48	-10.81	-	-6.83	-13.07	-
	Kharkiv	+2.95	+2.03	-	-2.09	-13.07	-	-33.44	-14.95	-	-6.18	-11.81	-
	Donetsk	+2.49	+2.03	-	-6.78	-12.03	-	-8.50	+9.49	-	-6.43	-12.23	-
	Kherson	+2.47	+1.49	-	-4.97	-4.97	-	-14.65	-11.50	-	-4.29	-10.79	-
	Lviv	+2.82	+1.01	-	+3.88	-13.93	-	+18.98	+39.90	-	-5.90	-11.69	-
Average of Absolute Changes		2.70	1.64		5.66	13.75		19.81	17.33		5.92	11.92	

The concentrations of O₃ all over Ukraine experienced similar increments (i.e., between 1% and 3%). It is also immediate that 2019 has been less polluted compared to 2021 in terms of O₃. This incremental behavior disaccords the subtractive pattern of the other three pollutants (excluding Lviv in terms of NO₂ and HCHO). Confirming these results, previous studies have revealed the increase of O₃ during exceptional events on both global and local scales (e.g., during a pandemic and political protests) [67]. The reason can be sought in the fact that the presence of this pollutant is a function of meteorological variables and precursors (e.g., NO_x and VOCs), and even in similar anthropogenic situations, the O₃ concentration can vary due to seasonal meteorological changes [115].

Satellite measurements of the NO₂ concentration are well-accepted proxies to address rapid changes in emissions from anthropogenic activities [116]. Setting Lviv aside, all other cities have reported significant reductions (up to about 25%) in NO₂ emissions in 2022. Kiev is the capital and Kharkiv, Donetsk, and Kherson are located in eastern Ukraine, which is the most industrialized territory of the country. Therefore, the observed reduction can attribute to diminished human activities. These results are well correlated with previous studies such as Zhang, Hu [66] who reported between 10.7% and 27.3% reduction in the NO₂ concentration in most Ukrainian cities. Taking another look at Figure 8 and Figure 9 reveals NO₂ spikes in eastern Ukraine (February 2022 around Kharkiv) that can attribute to multiple martial operations (e.g., combustions and explosions) on the war front (see Figure 3). A distinguished difference for 2022 vs. 2019 can be seen for Lviv which is characterized by 3.88% increase in the NO₂ concentration. As shown in Figure 5, Lviv lies 550 km west of the capital, which is not as, neither involved with the war nor industrialized, as the eastern part. Moreover, the values for 2022 vs. 2021 indicate a considerably larger difference compared to 2022 vs. 2019, meaning that 2021 has been the most polluted year. A potential reason for that could be the retrieval of anthropogenic activities after removing sanitary emergencies that were established in 2020. Another appreciable point is the higher variations of NO₂ in Kiev (i.e., -10.56% and -24.73%) compared to other cities. In addition to the reduced industrial functionality, the major decline in the population of this city can explain the alleviation of urban human circulation, and consequently, NO₂ emissions. The proportional relationship between NO₂ emission and population statistics has been well-professed before [117]. According to Leasure, Kashyap [118], only the initial evacuations that took place between February 24 and March 14 decline the population of Kiev by around two-thirds. Similarly, above 943000 people left the Kharkivska Oblast.

Analyzing the HCHO results denotes that, first, this pollutant changed the most among the considered gaseous pollutants. According to Popov, Iatsyshyn [4], the excessive concentration of

HCHO in Kiev (4.3 times more than the maximum permissible daily average) has been an issue since the past. Second, Lviv has reported again differently from other cities. In other words, the war has led to the reduction of the HCHO concentration in all places except Lviv (i.e., +18.98% for 2022 vs. 2019 and +39.90% for 2022 vs. 2021) along with Donetsk (+9.49% for 2022 vs. 2021). The most tangible changes for this pollutant are demonstrated in Kharkiv and Lviv having -33.44% and +39.90%, respectively. Earlier, non-natural origins of the HCHO emissions were explained as traffic and petrochemical industries, whose essential decline due to population displacement and industrial degradations can justify the significant reduction of the HCHO emission.

Based on the CO values, it can be derived that the war has caused widespread reductions of this pollutant. Similar to O₃, the changes in CO have been more or less gentle when comparing the five cities (i.e., between 4% and 7% for 2022 vs. 2019 and between 10% and 14% for 2022 vs. 2021). However, analogous to NO₂, Kiev is affected the most. Larger reductions that are calculated for 2022 vs. 2021 infer the highest concentration of CO in 2021. On the other hand, 2022 records indicate the lowest emissions which are expected by virtue of the war causing people evacuations and industrial degradations. As explained earlier, human activities such as fossil fuel consumption and waste incineration are responsible for around 60% of CO emissions. The considerable decline in the CO concentration during exceptional events has been also observed in previous studies (e.g., in Milan, Italy [119] and São Paulo, Brazil [120]) that investigated the effects of COVID-19 lockdowns [121].

At the end, another statistical analysis called t-test is carried out to demonstrate whether the changes in the air quality of Ukraine are statistically significant. The pivotal outcome of the t-test is p-value that when it is below 0.05 confirms a significant difference between the compared pairs [122]. For each pollutant and in each city, two comparisons are implemented, one for 2019 vs. 2022 and another for 2021 vs. 2022. The results are presented in Table 8.

Table 8: T-test results (p-values) for comparing the pollutant concentrations in 2022 vs BAU years

Area	O ₃		NO ₂		HCHO		CO	
	2019 vs. 2022	2021 vs. 2022	2019 vs. 2022	2021 vs. 2022	2019 vs. 2022	2021 vs. 2022	2019 vs. 2022	2021 vs. 2022
Kiev	0.0539	0.4250	0.2697	0.0087	0.2372	0.6362	0.0000	0.0001
Kharkiv	0.0431	0.3854	0.7414	0.0496	0.0273	0.2855	0.0015	0.0003
Donetsk	0.0331	0.2748	0.2032	0.0248	0.2255	0.8891	0.0002	0.0005
Kherson	0.0183	0.3508	0.2855	0.0411	0.3977	0.3447	0.0011	0.0004
Lviv	0.0305	0.5203	0.4101	0.0086	0.5158	0.0270	0.0011	0.0001

According to this table, there are various levels of significance for each pollutant. For instance, the O₃ records of 2019 vs. 2022 are significantly different because their p-values (with a small tolerance for Kiev) are below 0.05, while those calculated for 2021 vs. 2022 were higher than this threshold. In accordance with these results, Table 7 earlier reported a higher difference between 2019 vs. 2022 compared to 2021 vs. 2022. On the contrary, the p-values corresponding to NO₂ for 2021 vs. 2022 are less than 0.05 which indicate a statistically significant difference, while those calculated for 2019 vs. 2022 are considerably higher. Likewise, it was expected based on what presented in Table 7 in this regard. About the HCHO, only two p-values below 0.05 were obtained for Kharkiv in 2019 vs. 2022 and Lviv in 2021 vs. 2022 which respectively correspond to major changes -33.44% and +39.90 in Table 7. The smallest p-values were obtained for the CO (all p-values below 0.05) which infer the dynamics impacts of the war on this pollutant. From a comparative point of view, the CO results, too, can support the outcomes of Table 7, meaning, the lower the p-value is, the bigger the change was.

3.2.4 Further Discussion

Having a conclusive overview of the results suggests that this phase of the project could conduct proper AQM over Ukraine regarding the ongoing war crisis with Russia. Analyzing the emission of four gaseous pollutants including O₃, NO₂, HCHO, and CO based on Sentinel-5P data revealed both consistent and heterogeneous changes in the pattern of these pollutants in different parts of the country. Nevertheless, the findings altogether demonstrated the evident impact of the war on air pollution.

Generally speaking, there are key barriers to the way of proper acquisition, processing, and interpretation of observational data. In this sense, technical skills (e.g., coding in specific programming languages, statistical interpretations, etc.) are necessary for the data end-users [8]. Going deeper into the AQM using Sentinel-5P, it can be a problematic task due to issues such as a small number of daily observations and rapid diffusion of gaseous pollutants in the atmosphere [82]. With these issues in mind (e.g., data quality and completeness), the results may be subjected to uncertainties [45]. As observed in this phase, inconsistent records of the CH₄, and SO₂ led to their removal from assessments. Besides, the HCHO observations can be argued as far as the quality (e.g., negative concentrations) and completeness are concerned.

Although this phase was associated with the above limitations in the used data, it provided notable improvements. So far, and to the best knowledge of the author, the only studies that particularly dealt with the same subject are carried out by [Zhang, Hu \[66\]](#) and [Zalakeviciute, Mejia \[67\]](#). Along with the valuable findings, these works had weaknesses, too. For instance, [Zhang, Hu \[66\]](#)

investigated the effect of NO₂ only, and [Zalakeviciute, Mejia \[67\]](#), despite using a good dataset, bounded their research to the initial days of the war. In addition, they employed an averaged single observation to represent the whole of Ukraine and focused only on Kiev for more particular analyses. Hence, in this phase, it was also tried to cope with the shortcomings of previous literature toward creating a more comprehensive AQM to better manage the war-induced air quality issues in Ukraine.

4. PHASE 2: AIR QUALITY PREDICTION USING MACHINE LEARNING

After the AQM presented in the first part of the thesis, Phase 2 addresses the application of ML techniques for predicting air pollution during the war. For this purpose, a reliable dataset is provided and exposed to well-known ML models in order to predict the concentration of particulate matter 2.5 (PM_{2.5}). PM_{2.5} signifies the fraction of aerosol particles having an aerodynamic diameter less than 2.5 μm [\[123\]](#). These fine inhalable particles are mainly generated by domestic fuel combustion and urban traffic and can cause respiratory and cardiovascular health issues [\[124\]](#). Therefore, having a dependable estimation of its exposure can be of great importance, especially in vital situations such as military conflicts.

Due to the obligation of using ground measurements for acquiring PM_{2.5} records, air quality prediction in this section is carried out only for the capital city, i.e., Kiev. This matter is explained more clearly in the relevant section.

Similar to Phase 1, first the materials and methods are described. Next, the results are exhibited and associated with relevant discussion.

4.1 Materials and Methods

This section is overall concerned with three major parts. First, the used data is explained in terms of acquisition and preprocessing. Second, the methodology of the applied algorithms is explained. Lastly, this section defines the criteria that are considered for assessing the goodness of results.

4.1.1 Data Provision and Analysis

4.1.1.1 Parameter Selection

Preparing a reliable dataset is a prerequisite for executing an ML model. The dataset consists of sufficient homogeneous records of parameters that are divided into two categories: (i) one (or more) parameter(s) that is planned to be predicted called target parameter(s), and (ii) some factors that logically influence the target parameter called input factors.

In this study, the $PM_{2.5}$ concentration plays the role of a target parameter. Input factors, however, are composed of three major groups, namely meteorological factors, atmospheric factors, and temporal factors. The ML model explores the relationship between the target parameter and input factors i.e., the dependency of the $PM_{2.5}$ concentration on input factors. The model herewith learns how the target is changed with the variation of inputs, and it will be ready to predict the $PM_{2.5}$ concentration for new conditions.

Regarding the input factors, they are carefully selected with reference to previous studies [45, 46, 53, 125]. The selected parameters are classified as follows:

- (i) Meteorological factors: the critical influence of meteorological conditions on air pollution is evident. Seven factors including humidity (H), temperature (T), dew point (DP), wind speed (WSp), wind direction (WD), solar irradiance (SI), and sea level pressure (SLP) are selected.
- (ii) Atmospheric factors: these factors include the concentration of O_3 , NO_2 , and CO. Note that, due to the inconsistencies of the HCHO records, and also the presence of negative values, this factor was omitted in this phase.
- (iii) Temporal factors: by having temporal factors, the model is able to establish a connection between time and pollution. Since daily pollution is predicted here, the day number (DN) is chosen to represent temporal dependency.

4.1.1.2 Sources

Three different sources are used to build the dataset. These sources are described as follows:

- (i) PurpleAir website: daily concentrations of $PM_{2.5}$, as well as the corresponding H, T, DP, and SLP, are taken from the freely available records through the PurpleAir portal [126]. As Figure 16 illustrates, for Ukraine, there is a total of four registered ground stations near Kiev. Out of these stations, the records of both Hlepcha and Shepeleva sensors were suitable for this work, but due to the proximity of Shepeleva station (Longitude = 30.432216 and Latitude = 50.420316) to the central Kiev, the required $PM_{2.5}$ concentrations and meteorological data are downloaded from this station.
- (ii) The GEE: daily concentrations of the O_3 , NO_2 , and CO are downloaded from the Sentinel-5P records using the GEE platform.

- (iii) Solcast website: hourly records of three meteorological parameters, namely WSp, WD, and SI are taken from the Solcast portal [127]. These parameters are then converted into a daily format to fit the rest of the dataset.

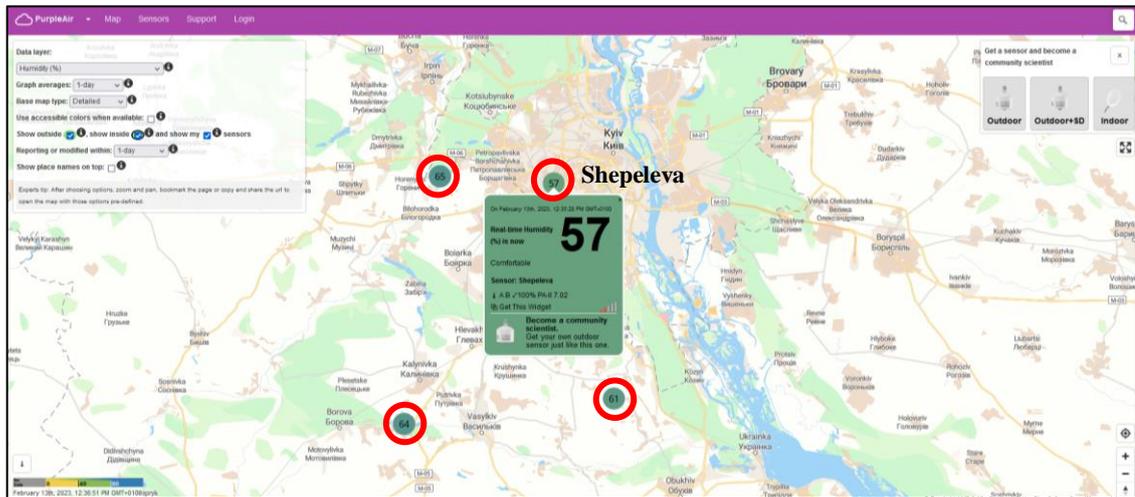


Figure 16: A view of the PurpleAir website showing registered ground stations near Kiev [126]

Table 9 gives information regarding the parameters of the dataset.

Table 9: Description of the used parameters

Type	Label	Source	Factor	Unit
Inputs	Meteorological	PurpleAir	Humidity (H)	(%)
			Temperature (T)	(°C)
			Dew Point (DP)	(°C)
			Sea Level Pressure (SLP)	(mbar)
			O ₃ concentration (O ₃)	($\mu\text{mol m}^{-2}$)
	Atmospheric	GEE	NO ₂ concentration (NO ₂)	($\mu\text{mol m}^{-2}$)
			CO concentration (CO)	($\mu\text{mol m}^{-2}$)
			Wind Speed (WSp)	(m s^{-1})
	Meteorological	Solcast	Wind Direction (WD)	(°)
			Solar Irradiance (SI)	(W m^{-2})
Targets	Temporal	Calendar	Day Number (DN)	-
	Atmospheric	PurpleAir	PM _{2.5} concentration (PM _{2.5})	($\mu\text{g m}^{-3}$)

4.1.1.3 Data Preprocessing and Splitting

Data preprocessing steps that are carried out to reach a suitable dataset are as follows:

- (i) Two steps are applied to the Solcast data. First, since the observations are recorded based on the coordinated universal time (UTC), the hours are corrected for the

Ukraine local time. Next, the hourly records of WSp, WD, and SI are averaged for each day to be compatible with other daily input factors. For this purpose, the Pivot Table feature in Excel is used.

- (ii) Shepeleva sensor uses two channels (named A and B) for measuring PM_{2.5}. Based on some recent studies that examined the quality of PurpleAir PM_{2.5} data, this data is recommended to be subjected to a correction before use. Equation (1) describes the applied correction [67, 128].

$$\text{Corrected } PM_{2.5} = 0.52 \times AV_{A\&B} - 0.086 \times H + 5.75, \quad (1)$$

where $AV_{A\&B}$ signifies the averaged PM_{2.5} for channels A and B. Also, H stands for relative humidity.

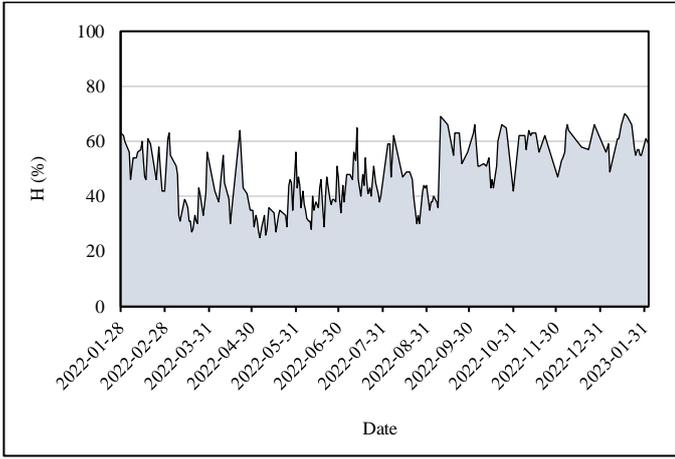
- (iii) Most input factors had some missing records throughout the dataset. Thus, the corresponding records were eliminated.

Eventually, the prepared dataset consists of 212 daily records in the period 27 January 2022 to 06 February 2023. Given the number of included parameters (11 input factors + 1 target parameter), a 212×12 Excel file is generated.

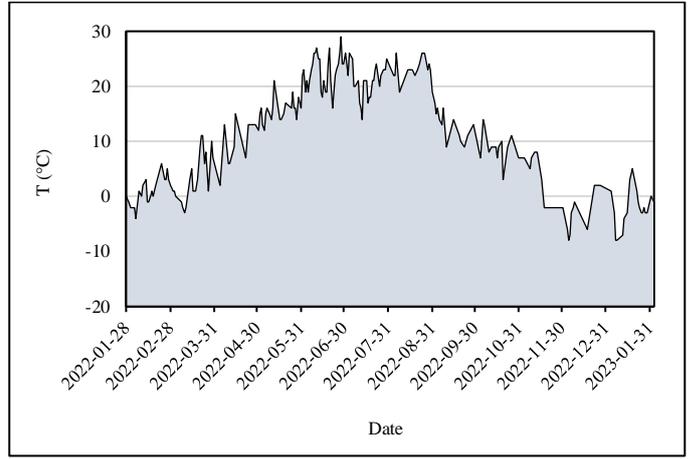
In an ML-based implementation, the dataset is split into two different sub-sets, namely training data and testing data. As the name suggests, the first fraction is used for training the models. During this process, the relationship between the target and inputs is analyzed and learned by the model. Since the second fraction (i.e., testing data) has not been confronted by the model, it is used after training for evaluating the generalizability of the model. In other words, the performance of the model in dealing with unseen conditions is assessed using testing data [129, 130]. For this study, 80% of data (i.e., 170 daily records) are randomly chosen to form the training dataset and the remaining 20% (i.e., 42 daily records) form the testing dataset.

4.1.1.4 Description and Statistics

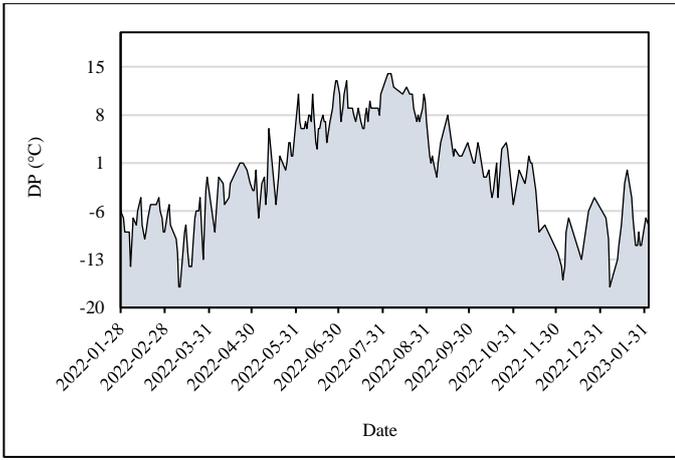
The statistical description of the dataset gives a clear vision of the data conditions. In Figure 17, the variation of the input factors and PM_{2.5} in the period of interest is depicted. Concerning the gaseous pollutants, it is seen that the O₃ shows a decreasing behavior moving toward the end of 2022, whereas the NO₂ peaks several times in late 2022 and early 2023. Moreover, evaluating the PM_{2.5} trend demonstrates sharp increments with the beginning of the war in February and March 2022, as well as the last three months of 2022.



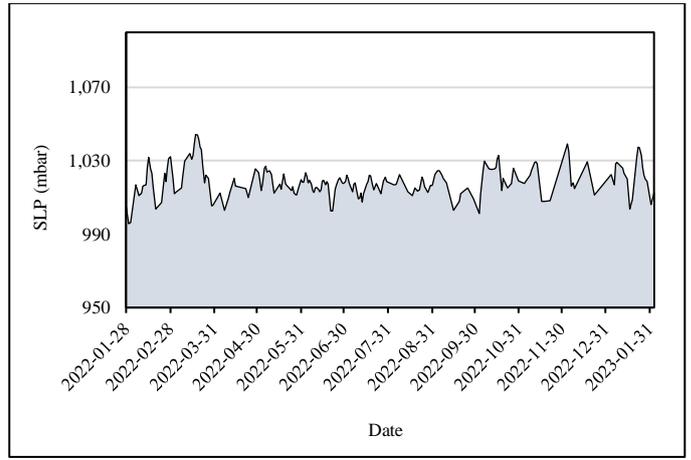
(a)



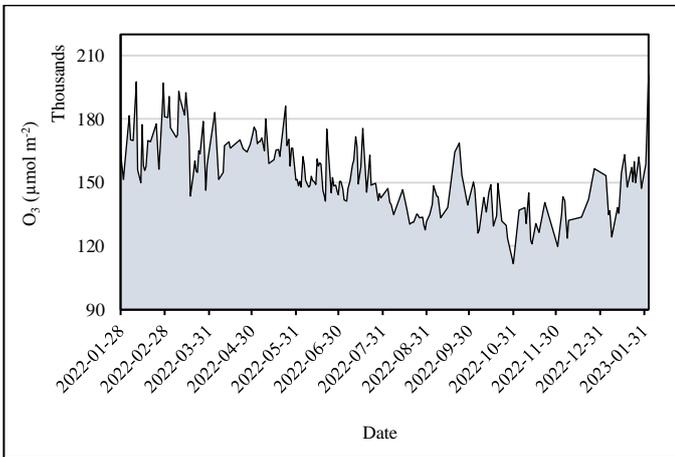
(b)



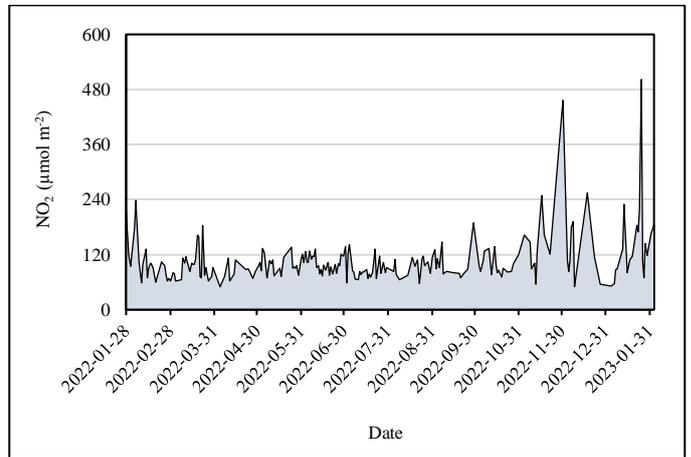
(c)



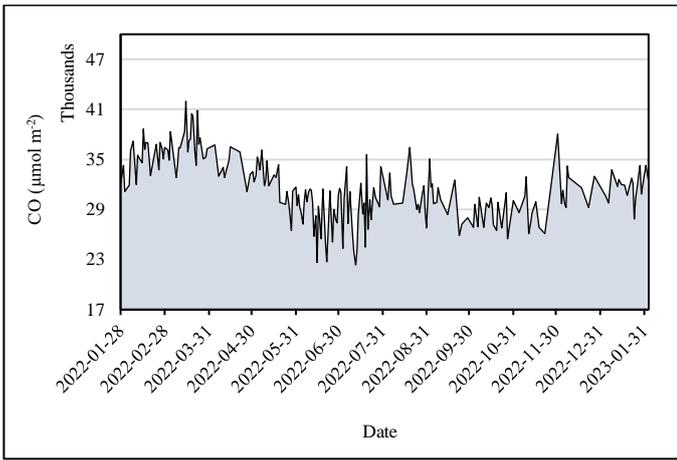
(d)



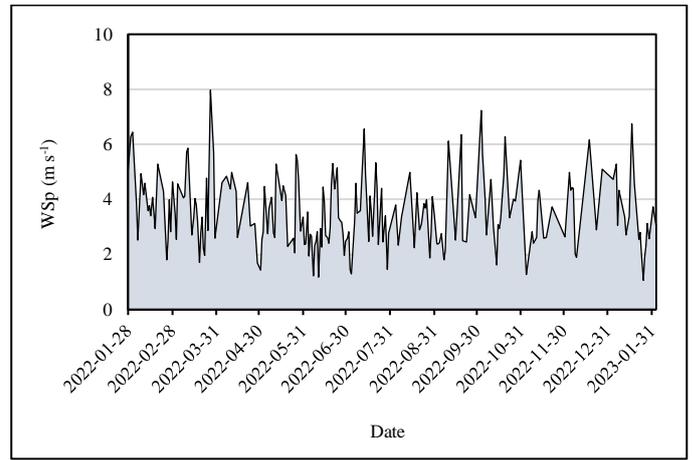
(e)



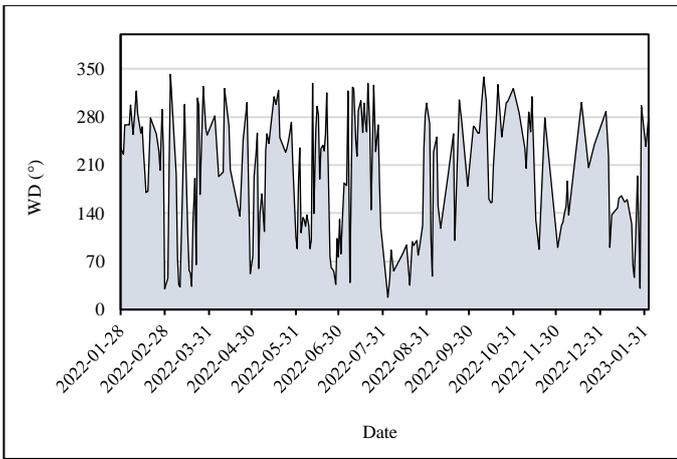
(f)



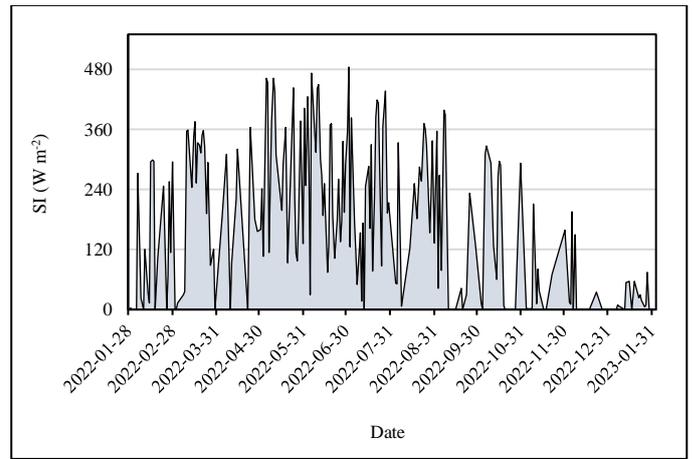
(g)



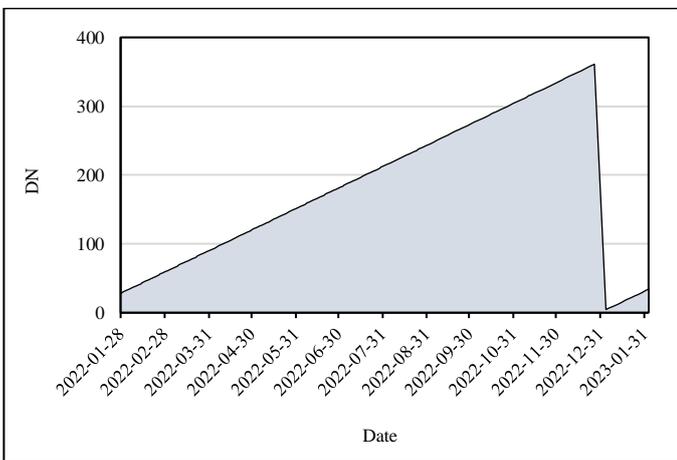
(h)



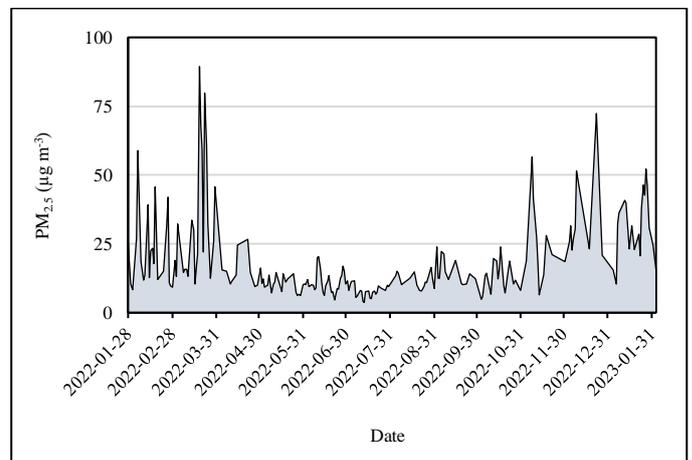
(i)



(j)



(k)



(l)

Figure 17: Variation of the dataset parameters with time

Using Excel, descriptive statistics of the dataset are produced and the values of mean, standard deviation, sample variance, skewness, minimum, and maximum are given in Table 10. The concentration of $PM_{2.5}$ ranges from 3.59 to $89.42 \mu\text{g m}^{-3}$ with an average of $18.08 \mu\text{g m}^{-3}$.

Table 10: Descriptive statistics of the dataset

Factor	Mean	Standard Deviation	Sample Variance	Skewness	Minimum	Maximum
H (%)	46.87	11.48	131.75	0.08	25.00	70.00
T (°C)	11.18	9.84	96.92	-0.11	-8.00	29.00
DP (°C)	-0.24	7.93	62.88	-0.13	-17.00	14.00
SLP (mbar)	1018.89	8.45	71.37	0.38	995.80	1044.30
O ₃ (μmol m ⁻²)	153546.03	16786.08	281772492.90	0.31	111561.71	200901.94
NO ₂ (μmol m ⁻²)	106.31	51.54	2656.15	4.16	49.50	501.80
CO (μmol m ⁻²)	31476.53	3679.24	13536798.57	0.11	22306.92	41959.29
WSp (m s ⁻¹)	3.52	1.28	1.64	0.63	1.05	7.97
WD (°)	194.58	88.51	7834.61	-0.26	17.88	341.96
SI (W m ⁻²)	177.42	146.92	21586.53	0.27	0.00	484.58
DN	163.56	95.67	9152.73	0.16	4.00	361.00
PM _{2.5} (μg m ⁻³)	18.08	14.16	200.60	2.27	3.59	89.42

Moreover, Table 11 contains the results of the correlation assessment among all parameters. The reported values can vary in [-1, 1] so that -1 means the two parameters are completely adversely proportional, and vice versa, 1 means the two parameters are completely directly proportional. The correlation of PM_{2.5} with the input factors is given in the last row. The largest values are obtained for CO (0.44) and SLP (0.32), while the smallest values are calculated for the two temperature factors (i.e., T and DP with correlations of -0.49 and -0.50, respectively).

Table 11: Correlation analysis of the parameters

	H	T	DP	SLP	O ₃	NO ₂	CO	WSp	WD	SI	DN	PM _{2.5}
H	1.00											
T	-0.50	1.00										
DP	-0.14	0.92	1.00									
SLP	-0.23	-0.26	-0.39	1.00								
O ₃	-0.22	-0.18	-0.30	-0.04	1.00							
NO ₂	0.12	-0.25	-0.24	0.18	-0.19	1.00						
CO	-0.18	-0.39	-0.52	0.28	0.35	0.08	1.00					
WSp	0.22	-0.23	-0.16	-0.30	0.12	-0.23	-0.04	1.00				
WD	0.14	-0.08	-0.03	-0.31	0.12	-0.14	-0.15	0.31	1.00			
SI	-0.78	0.53	0.26	0.20	0.06	-0.07	0.12	-0.23	-0.13	1.00		
DN	0.12	0.31	0.40	-0.05	-0.61	0.01	-0.49	-0.04	0.06	0.02	1.00	
PM _{2.5}	0.15	-0.49	-0.50	0.32	-0.05	0.30	0.44	-0.26	-0.09	-0.18	-0.22	1.00

4.1.2 Artificial Intelligence

MATLAB R2017b is used for implementing all artificial intelligence algorithms. In the following, the used techniques are described.

4.1.2.1 Artificial Neural Network (ANN)

The ANN is distinguished as a sophisticated ML method that works based on the function of neural network in the human brain. The basic idea of neural learning was first proposed by [McCulloch and Pitts \[131\]](#). So far, different types of ANNs have gained huge popularity, due to their ability of efficient non-linear analysis. Multi-layer perceptron neural network (MLPNN) is a potential class of ANN that, as the name implies, has a layered structure. Many studies have proven that even a simply structured MLPNN can be a universal approximator [\[132\]](#). Concerning specific applications, authors have broadly used the MLPNN for predicting various air quality parameters (e.g., [Ameer, Shah \[133\]](#), [Liu and Guo \[134\]](#), [Suresh, Kiranmayee \[135\]](#), etc.)

In this model, data is transmitted forward via three kinds of layers, namely input, hidden, and output layers. As Figure 18 displays, each layer can hold a number of neurons. A three-layered network with n input neurons, m hidden neurons, and z output neurons is signified as MLPNN (n, m, z). The neurons throughout the network are connected by weights. Also, one bias parameter is specified for hidden and output neurons. These two groups of neurons (i.e., hidden and output) employ linear/non-linear activation functions in order to perform the calculations.

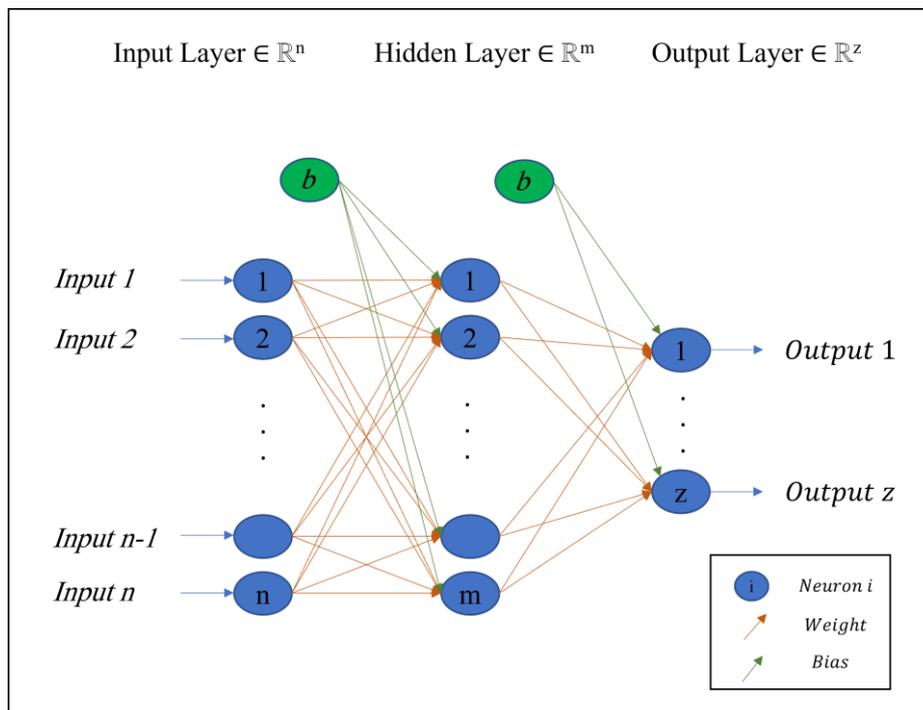


Figure 18: The topology of a general MLPNN(n, m, z)

The weights and biases of the MLPNN are tuned by specific algorithms in order to train the model. Equation (2) describes the mathematical relationship between the inputs (x_i), outputs (y_k), weights, and biases.

$$y_k = f(\sum_j w_{jk} \times f(\sum_i w_{ij} \times x_i + b_j) + b_k), \quad (2)$$

where w_{ij} represent the weights that connect the input and hidden neurons along with the corresponding bias b_j and w_{jk} represent the weights that connect the hidden and output neurons along with the corresponding bias b_k [136]. Employing two so-called strategies feed-forward and back-propagation, the MLPNN first process the data forward, and after producing the output, the error is calculated, then, the network propagates backward to adjust the parameters with the objective of reducing the error [137].

4.1.2.2 Electromagnetic Field Optimization (EFO)

[Abedinpourshotorban, Shamsuddin \[138\]](#) designed the EFO algorithm as one of the physical-inspired metaheuristic algorithms. This algorithm has been earlier used for optimizing MLPNN for purposes like groundwater level modeling [139] and solar energy estimation [140]. The population in EFO is represented by a set of so-called members electromagnetic particles (EP). As shown in Figure 19, there can be three polarities, namely positive, neutral, and negative, for the EPs.

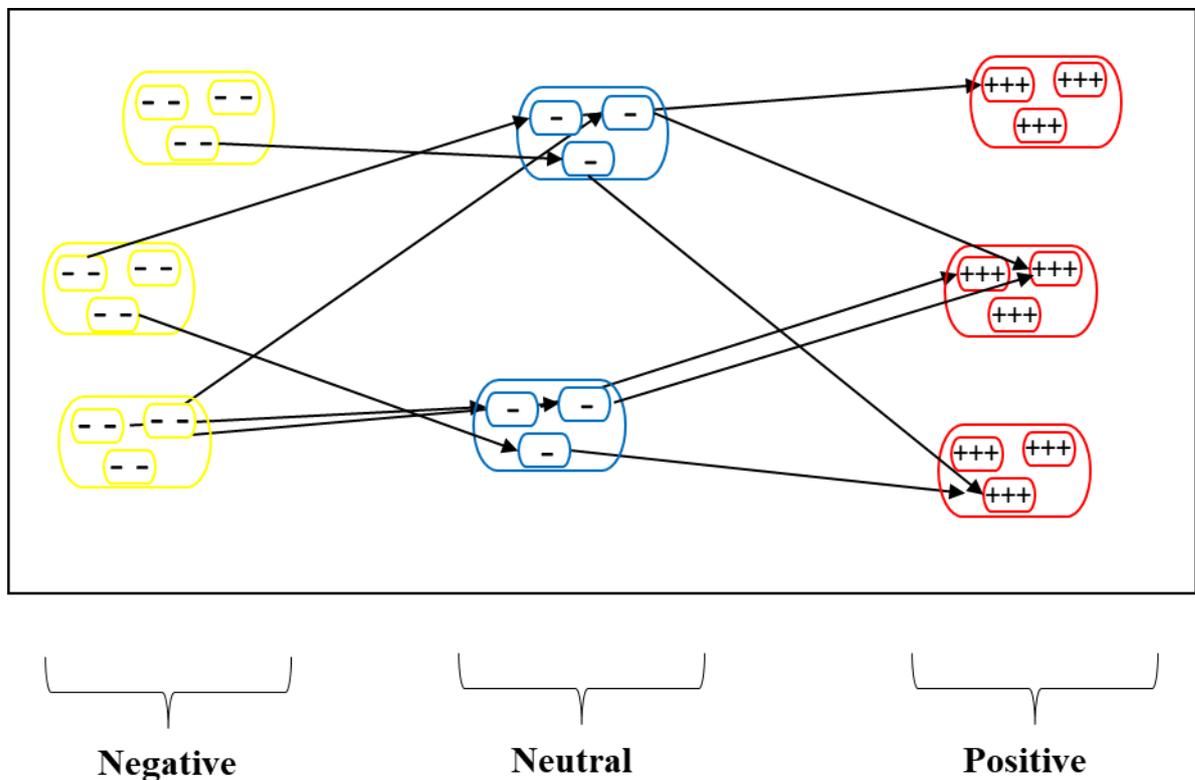


Figure 19: Three polarities and forces among the EPs

For optimization, the EFO population considers the repulsion-attraction interaction to update the population (i.e., produce new EPs) within the created electromagnetic field. The mathematical description is as follows:

- (i) Population Generation: Based on Equation (3), the particles are randomly generated in the search space bounded within the lower bound (LB_j) and upper bound (UB_j).

$$EP_j^i = LB_j + rnd(UB_j - LB_j), i = 1, 2, \dots, S_{Pop} \text{ and } j = 1, 2, \dots, D \quad (3)$$

where rnd symbolizes a random value in $[0, 1]$, S_{Pop} stands for the population size, and D represents the dimension. The population is next sorted with respect to the fitness order of the EPs.

- (ii) Classification: The population is grouped into three positive (P), neutral (Nt), and negative (N) fields. In this regard, the best-fitted EPs go into the positive field with a population ratio P_{field} , the worst-fitted EPs go into the negative field with a population ratio N_{field} , and the remaining EPs go into the neutral field.
- (iii) New EP Generation: One random particle is selected from each field to incorporate based on Equation (6) for generating a new EP.

$$rEP_j^{New} = \begin{cases} rEP_j^{Pj}, & rnd < Ps_{rate} \\ rEP_j^{Nt} + (\sigma \times rnd)(rEP_j^P - rEP_j^{Nt}) - rnd(rEP_j^N - rEP_j^{Nt}), & otherwise \end{cases} \quad (4)$$

where particles selected from the P, N, and Nt fields are represented by rEP^P , rEP^N , and rEP^{Nt} , respectively. σ is the golden ratio = 1.618 and Ps_{rate} stands for the probability of choosing variables of the new EP from the P field.

- (iv) Randomization: One electromagnet of the generated EP goes through a randomization process based on Equation (5) where irt denotes an integer iteration counter.

$$EP_{itr}^{New} = LB_{itr} + rnd(UB_{itr} - LB_{itr}), \text{ if } rnd > R_{rate} \quad (5)$$

in which, R_{rate} gives the probability of changing one electromagnet of the EP with a random electromagnet.

- (v) Selection: In this process, the fitness of the new EP is assessed and if it has higher fitness compared to the worst existing EP, the new EP replaces the worst one [141].

4.1.3 Accuracy Evaluation Method

There are various statistical indicators that can reflect the accuracy of predictions. In order to evaluate the performance of the ML models, this study uses three well-known indicators, namely root mean square error (RMSE), mean absolute error (MAE), and Pearson correlation coefficient

(R_P). While the RMSE and MAE calculate the error of prediction i.e., the difference between the output and target $PM_{2.5}$, the R_P indicates the correlation between these values. Relatively speaking, a smaller RMSE and MAE represent a better prediction, while R_P ranges from 0 to 1 so that 1 reflects the largest accuracy and vice versa.

Given $PM_{2.5i_t}$ and $PM_{2.5i_o}$ as the target and output $PM_{2.5}$ with respective average values of $\overline{PM_{2.5i_t}}$ and $\overline{PM_{2.5i_o}}$, Equations (6) and (7) formulate the RMSE and MAE, and Equation (8) formulates the R_P .

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (PM_{2.5i_t} - PM_{2.5i_o})^2}, \quad (6)$$

$$MAE = \frac{1}{N} \sum_{i=1}^N |PM_{2.5i_t} - PM_{2.5i_o}|, \quad (7)$$

$$R_P = \frac{\sum_{i=1}^N (PM_{2.5i_o} - \overline{PM_{2.5i_o}})(PM_{2.5i_t} - \overline{PM_{2.5i_t}})}{\sqrt{\sum_{i=1}^N (PM_{2.5i_o} - \overline{PM_{2.5i_o}})^2} \sqrt{\sum_{i=1}^N (PM_{2.5i_t} - \overline{PM_{2.5i_t}})^2}}, \quad (8)$$

where N stands for the number of data.

4.1.4 Scenarios

The main objective of defining scenarios is to try several viable conditions to attain the most optimum prediction. Regarding the used dataset, four scenarios are defined for the prediction of air pollution. Two scenarios are about the temporal aspect of the atmospheric data. In one dataset, the $PM_{2.5}$ in the day i is considered as the function of input factors recorded for the day i (called dataset i), while in the other dataset, the $PM_{2.5}$ in the day i is considered as the function of input factors in day $i-1$ (called dataset $i-1$). [142, 143]

Two scenarios are also defined based on the dimension of the problem. The first dataset is in the original form and comprises 11 input factors (called original-input dataset), while the second dataset comprises a lower number of inputs (called input-reduced dataset) with reference to feature selection analysis. [144]

The scenarios are summarized as follows:

- (i) Scenario 1: dataset i & original-input dataset
- (ii) Scenario 2: dataset $i-1$ & original-input dataset
- (iii) Scenario 3: dataset i & input-reduced dataset
- (iv) Scenario 4: dataset $i-1$ & input-reduced dataset

4.2 Results and Discussion

4.2.1 Feature Selection Using Principal Component Analysis (PCA)

Due to the large number of input factors involved, the ML model has to confront a high-dimensional problem. It has been proven that trimming the dimension of the problem is a viable idea for reducing the computational cost, as well as increasing the accuracy occasionally. In this regard, there are several statistical methods that point out the importance of the input factors. The factors with the lowest importance can be eliminated, and then check if it improves the accuracy. In this study, a well-known technique, namely principal component analysis (PCA) is employed. The basis of the PCA is transferring the existing interrelated variables into some orthogonal (independent) variables called principal components (PCs). Each PC contains a linear combination of the original variables and is uncorrelated to the other PCs [145]. The PCA analyses are carried out in the IBM SPSS Statistics 22 environment.

Figure 20 shows the scree plot of the PCA results. This plot depicts the eigenvalues for the constructed PCs. Although the cutoff point is observed around the fourth PC, only the PCs with eigenvalues greater than 1 are considered significant [146]. The first three PCs have eigenvalues 3.062, 2.435, and 1.804. Based on the detailed results in Table 12, these three PCs account for around 66.37% variance in data.

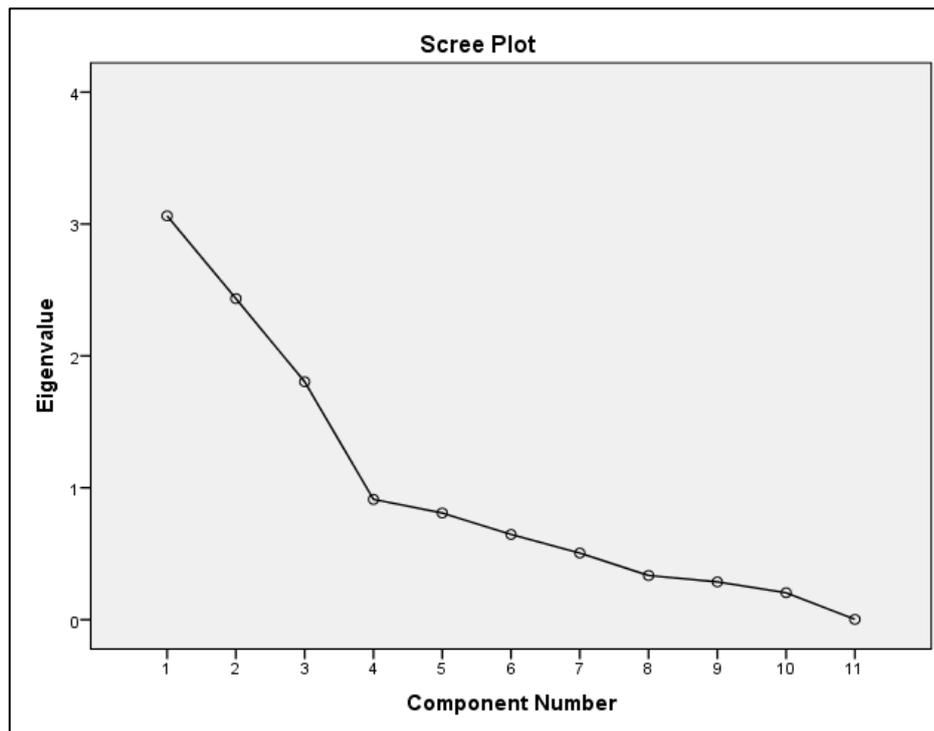


Figure 20: Scree plot of the PCA analysis

Table 12: Detailed variance analysis of the PCs

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	3.062	27.836	27.836	3.062	27.836	27.836	2.715	24.679	24.679
2	2.435	22.134	49.970	2.435	22.134	49.970	2.614	23.763	48.442
3	1.804	16.398	66.368	1.804	16.398	66.368	1.972	17.926	66.368
4	.912	8.290	74.658						
5	.809	7.354	82.012						
6	.646	5.877	87.889						
7	.505	4.592	92.480						
8	.335	3.044	95.525						
9	.286	2.601	98.126						
10	.204	1.855	99.981						
11	.002	.019	100.000						

In the following, three rotation methods, namely Varimax, Quartimax, and Equamax are applied to the selected PCs in order to identify the most important input factors. In this sense, Kaiser Criterion [147] is used to decide which factors should be retained and which should be disregarded. The results are given in Table 13. Referring to previous studies [144, 148], 0.75 can be considered a threshold of factor loading to determine strong factors. Input factors having absolute factor loadings larger than (or equal to) 0.75 are highlighted in Table 13 (negatively and positively strong loadings are highlighted in green and red, respectively).

Table 13: Rotated factor loadings and communalities for the used rotation methods

Input	Rotated Component Matrix (Varimax)			Rotated Component Matrix (Quartimax)			Rotated Component Matrix (Equamax)		
	PC 1	PC 2	PC 3	PC 1	PC 2	PC 3	PC 1	PC 2	PC 3
H	0.239	-0.856	0.166	0.233	-0.857	0.170	0.242	-0.856	0.165
T	0.535	0.796	0.154	0.540	0.793	0.149	0.532	0.797	0.156
DP	0.711	0.538	0.253	0.715	0.534	0.250	0.709	0.539	0.255
SLP	-0.274	0.029	-0.725	-0.275	0.028	-0.724	-0.273	0.030	-0.725
O ₃	-0.751	0.182	0.335	-0.750	0.188	0.335	-0.752	0.179	0.335
NO ₂	0.046	-0.315	-0.578	0.043	-0.318	-0.577	0.047	-0.314	-0.579
CO	-0.750	0.019	-0.264	-0.750	0.023	-0.263	-0.750	0.017	-0.264
WSp	-0.145	-0.295	0.626	-0.146	-0.292	0.628	-0.144	-0.297	0.626
WD	-0.012	-0.166	0.610	-0.013	-0.163	0.611	-0.012	-0.167	0.610
SI	-0.076	0.842	-0.209	-0.071	0.841	-0.213	-0.079	0.842	-0.207
DN	0.797	-0.039	-0.084	0.796	-0.045	-0.085	0.797	-0.036	-0.084

As is seen, significant factors are obtained only from the first and second PCs, that together account for nearly 50% variance in data (see Table 12). However, all three models (i.e., Varimax,

Quartimax, and Equamax) agree on the retention of H, T, O₃, CO, SI, and DN, because these six factors satisfy the loading threshold [144, 149].

With reference to these results, the input-reduced dataset for implementing Scenarios 3 and 4 is created by retaining H, T, O₃, CO, SI, and DN and eliminating the other five factors (i.e., DP, SLP, NO₂, WSp, and WD).

4.2.2 Method Configuration and Scenario Evaluation

Determining the configuration of the ANN is an important step in such studies. The reason is that the prediction accuracy can be affected by the computational architecture of the model. Despite some of the characteristics that should be decided by trial-and-error, some others can be satisfactorily determined from experience and previous attempts. For instance, as per Section 4.1.2.1, the applied MLPNN model can possess a minimum of three layers, but it has been broadly shown that a three-layered network can excellently model complicated problems [150]. Hence, one hidden layer is embedded in the used MLPNN. Table 14 addresses the characteristics that are fixed for all used MLPNNs (i.e., in all scenarios).

Table 14: Fixed characteristics of the applied MLPNNs

No. of Layers	No. of Neurons in Output Layer	Activation Function in Hidden Layer	Activation Function in Output Layer	Training Algorithm	Max. No. of Epochs
3	1	Tansig ($y = \frac{2}{1 + \exp(x)} - 1$)	Purelin ($x = y$)	Trainlm (Levenberg Marquardt) [151]	1000

As a characteristic of the MLPNN, the number of neurons in the input and output layer is determined by (equals) the number of inputs and outputs, respectively. Given the PM_{2.5} as the single output of all scenarios, one neuron lies in the output layer (see Table 14). This is while the number of inputs varies based on the ongoing scenario and accordingly equals the number of input neurons (i.e., eleven inputs in Scenarios 1 and 2 and six inputs in Scenarios 3 and 4). The number of neurons in the hidden layer is determined with respect to the data conditions. Due to imposing different conditions in each scenario, this number is determined by trial-and-error efforts. The performed process includes running the intended MLPNN with a variable number of neurons in the hidden layer (m). Ten values are examined so that $m = 1, 2, \dots, 10$, and the quality of the results is quantified by the RMSE criterion for the training and testing phases (TR and TS, respectively). Each model is run six times (i.e., Iterations 1-6) and their average RMSEs are calculated. In the

end, the configuration that yields the smallest RMSEs is chosen. The results are exhibited in the following tables with a color-intensity visualization.

Table 15: The results of trial-and-error process for finding the best m in Scenario 1

Model Number	m	MLPNN Result																
		RMSE																Total ranking
		Iteration 1		Iteration 2		Iteration 3		Iteration 4		Iteration 5		Iteration 6		Average		Average Ranking		
		TR1	TS1	TR2	TS2	TR3	TS3	TR4	TS4	TR5	TS5	TR6	TS6	TR	TS	TR	TS	
1	1	14.17	12.45	10.25	9.79	9.45	9.15	10.36	10.07	9.90	9.27	11.54	9.89	10.95	10.10	9	5	14
2	2	9.50	9.24	9.45	9.01	10.28	9.43	8.79	9.95	11.44	9.90	9.57	8.48	9.84	9.34	3	2	5
3	3	9.23	10.08	9.10	8.46	8.13	10.03	10.59	9.44	11.75	12.25	10.71	8.54	9.92	9.80	4	4	8
4	4	9.54	8.33	9.01	9.11	10.54	10.14	9.62	9.28	9.37	8.20	10.07	10.94	9.69	9.33	2	1	3
5	5	13.71	11.27	11.25	13.39	11.71	11.39	9.54	9.45	11.21	8.87	8.68	9.47	11.02	10.64	10	7	17
6	6	8.73	9.94	7.97	13.28	8.17	9.59	9.21	11.20	11.28	13.17	11.21	11.90	9.43	11.51	1	10	11
7	7	9.12	11.09	12.12	11.51	9.30	10.46	9.78	8.76	8.93	11.66	10.49	10.24	9.96	10.62	6	6	12
8	8	9.20	10.56	9.12	10.87	12.43	11.02	10.15	8.67	9.37	8.30	9.90	8.36	10.03	9.63	7	3	10
9	9	8.91	13.24	9.57	10.40	12.10	9.94	12.61	10.06	9.20	11.28	8.79	11.00	10.20	10.98	8	9	17
10	10	8.50	10.56	8.66	11.89	9.08	11.06	9.47	8.99	10.81	10.00	13.17	11.62	9.95	10.69	5	8	13

Table 16: The results of trial-and-error process for finding the best m in Scenario 2

Model Number	m	MLPNN Result																
		RMSE																Total ranking
		Iteration 1		Iteration 2		Iteration 3		Iteration 4		Iteration 5		Iteration 6		Average		Average Ranking		
		TR1	TS1	TR2	TS2	TR3	TS3	TR4	TS4	TR5	TS5	TR6	TS6	TR	TS	TR	TS	
1	1	9.55	13.04	10.83	11.69	10.47	11.10	11.73	14.05	11.61	11.81	10.80	10.98	10.83	12.11	9	4	13
2	2	10.47	11.34	11.15	14.10	10.53	10.44	10.21	10.23	9.66	11.45	9.96	12.72	10.33	11.71	6	2	8
3	3	9.50	9.83	9.86	10.13	10.51	12.06	10.21	11.76	11.06	11.82	10.43	10.82	10.26	11.07	4	1	5
4	4	12.06	13.73	10.47	12.36	9.28	10.78	12.89	14.43	9.97	11.02	8.66	12.09	10.56	12.40	7	6	13
5	5	8.93	12.59	12.06	12.41	10.77	12.23	9.83	13.15	11.13	13.55	11.56	12.97	10.71	12.81	8	8	16
6	6	9.76	10.13	10.36	11.79	11.81	13.75	8.95	10.97	12.19	14.20	8.78	11.71	10.31	12.09	5	3	8
7	7	11.37	15.31	8.69	14.13	9.61	12.96	10.57	12.09	8.88	13.29	8.13	12.76	9.54	13.42	1	10	11
8	8	10.40	13.70	12.35	14.46	10.02	11.21	11.56	14.01	10.61	13.62	10.86	11.72	10.97	13.12	10	9	19
9	9	9.43	15.40	9.72	12.83	10.57	11.54	9.53	10.55	10.92	12.42	10.93	11.88	10.18	12.44	3	7	10
10	10	9.98	14.92	9.78	12.48	9.22	12.32	11.17	11.16	9.94	12.03	9.35	10.47	9.91	12.23	2	5	7

Table 17: The results of trial-and-error process for finding the best m in Scenario 3

Model Number	m	MLPNN Result																
		RMSE																Total ranking
		Iteration 1		Iteration 2		Iteration 3		Iteration 4		Iteration 5		Iteration 6		Average		Average Ranking		
		TR1	TS1	TR2	TS2	TR3	TS3	TR4	TS4	TR5	TS5	TR6	TS6	TR	TS	TR	TS	
1	1	10.36	13.56	12.86	17.69	10.87	12.99	12.85	17.86	10.83	14.37	10.66	14.10	11.40	15.10	10	9	19
2	2	10.65	13.76	11.51	13.89	10.97	14.28	12.80	17.70	10.66	13.55	10.14	10.95	11.12	14.02	9	6	15
3	3	9.63	13.01	11.17	13.83	10.54	12.35	12.34	15.71	11.40	14.75	11.53	15.07	11.10	14.12	8	7	15
4	4	10.46	13.12	11.13	14.48	9.71	12.31	10.08	12.84	10.03	12.72	11.14	12.44	10.43	12.99	4	2	6
5	5	11.95	16.24	12.15	16.94	10.19	11.94	10.79	13.73	10.53	15.36	10.62	13.62	11.04	14.64	7	8	15
6	6	8.85	10.80	11.24	16.17	10.09	13.14	9.82	13.87	10.28	13.33	10.91	14.30	10.20	13.60	3	5	8
7	7	10.17	13.39	10.56	13.54	11.06	13.05	11.31	14.80	11.12	12.90	10.45	13.72	10.78	13.57	6	4	10
8	8	8.93	10.28	10.61	10.72	9.98	11.14	9.83	10.28	9.67	13.23	9.64	10.82	9.77	11.08	1	1	2
9	9	8.82	14.90	10.74	12.91	10.04	12.18	9.87	14.51	10.05	12.02	10.59	13.67	10.02	13.37	2	3	5
10	10	11.30	16.67	10.49	12.64	12.41	16.30	10.00	13.74	9.60	13.67	10.53	18.49	10.72	15.25	5	10	15

Table 18: The results of trial-and-error process for finding the best m in Scenario 4

Model Number	m	MLPNN Result																
		RMSE																Total ranking
		Iteration 1		Iteration 2		Iteration 3		Iteration 4		Iteration 5		Iteration 6		Average		Average Ranking		
		TR1	TS1	TR2	TS2	TR3	TS3	TR4	TS4	TR5	TS5	TR6	TS6	TR	TS	TR	TS	
1	1	14.15	14.41	11.42	7.52	11.09	12.33	11.04	9.14	9.76	14.42	11.05	11.59	11.42	11.57	9	7	16
2	2	10.35	11.50	12.10	6.68	10.21	9.24	13.94	11.85	12.09	16.79	10.57	12.13	11.55	11.36	10	6	16
3	3	11.36	12.06	11.80	6.53	11.00	11.49	11.00	9.68	9.48	11.75	9.11	13.06	10.62	10.76	6	2	8
4	4	10.62	11.40	9.88	10.21	10.37	8.20	10.99	8.56	10.69	14.57	11.01	13.15	10.59	11.01	4	3	7
5	5	11.55	11.39	14.40	9.64	12.64	14.23	9.22	8.48	9.01	12.07	10.57	12.20	11.23	11.33	8	5	13
6	6	9.15	13.46	17.58	12.25	11.57	11.20	9.64	11.49	9.37	14.82	9.92	12.34	11.21	12.59	7	10	17
7	7	9.52	9.93	10.64	8.86	12.14	14.03	10.42	9.92	10.77	15.49	10.21	12.46	10.62	11.78	5	8	13
8	8	9.73	9.56	9.20	6.81	10.31	9.82	11.55	10.00	9.68	13.92	9.74	11.11	10.04	10.20	1	1	2
9	9	9.80	11.33	9.69	9.00	10.38	11.53	13.58	10.84	9.48	12.97	8.63	11.63	10.26	11.22	3	4	7
10	10	10.90	11.48	10.08	9.77	9.29	11.08	10.46	12.98	9.78	14.81	9.77	11.76	10.05	11.98	2	9	11

According to Table 15, Table 16, Table 17, and Table 18, the best m (i.e., m with the smallest total ranking) for Scenario 1, Scenario 2, Scenario 3, and Scenario 4 is 4, 3, 8, and 8, respectively. Imposing these m values creates ANNs that are named MLPNN(11, 4, 1), MLPNN(11, 3, 1), MLPNN(6, 8, 1)*, and MLPNN(6, 8, 1), respectively used for Scenario 1, Scenario 2, Scenario 3, and Scenario 4.

Next, MLPNN(11, 4, 1), MLPNN(11, 3, 1), MLPNN(6, 8, 1)*, and MLPNN(6, 8, 1) each are implemented several times and the ones with outstanding accuracies are identified. Table 19 presents the results of these four models. Referring to the respective RMSEs 9.82, 9.89, 10.20, and 9.77 along with the MAEs 6.44, 6.96, 7.06, and 6.09, the MLPNN(6, 8, 1) outperforms other models in the training phase. In terms of R_p s 0.75, 0.72, 0.71, 0.75, MLPNN(11, 4, 1) and MLPNN(6, 8, 1) have the same accuracy, but higher than the two other models. In the testing phase, the superiority of MLPNN(11, 4, 1) is more evident than in the training phase. In this sense, RMSEs 8.75, 9.77, 10.51, and 7.51, MAEs 6.37, 7.05, 7.71, 5.92, as well as R_p s 0.62, 0.70, 0.67, and 0.79 indicate lower errors and higher correlation of the MLPNN(6, 8, 1) compared to other three models.

Table 19: The accuracy results obtained for four implemented scenarios

Scenario	Used Model	Train			Test		
		RMSE	MAE	R_p	RMSE	MAE	R_p
1	MLPNN(11, 4, 1)	9.82	6.44	0.75	8.75	6.37	0.62
2	MLPNN(11, 3, 1)	9.89	6.96	0.72	9.77	7.05	0.70
3	MLPNN(6, 8, 1)*	10.20	7.06	0.71	10.51	7.71	0.67
4	MLPNN(6, 8, 1)	9.77	6.09	0.75	7.51	5.92	0.79

All in all, the dataset designated to Scenario 4 yielded the most reliable analysis of air pollution. It means that the ideas exerted in Scenario 4 have been helpful for attaining a more desirable estimation of $PM_{2.5}$. More clearly, reducing the input dimension (from 11 to 6) using the PCA technique coupled with predicting one-day ahead $PM_{2.5}$ had positive effects on the performance of the MLPNN.

Therefore, the next evaluations are carried out using the dataset of Scenario 4 and the ANN model is represented by MLPNN(6, 8, 1).

4.2.3 Optimization and Improvement

Once the best approach (i.e., the most accurate scenario) is determined, the corresponding model is here optimized using the EFO metaheuristic algorithm (Section 4.1.2.2). The purpose of this

optimization is to see whether the accuracy of prediction can be further enhanced. The results of the optimized model are then compared to the benchmark models to validate its performance.

Figure 21 shows the topology of the model selected from Scenario 4, i.e., MLPNN(6, 8, 1). In this model, six input factors of the i - I^{th} day are analyzed to predict the $PM_{2.5}$ of the day i . Between the input and hidden layers, there are $6 \times 8 = 48$ weights and $1 \times 8 = 8$ biases. As for between the hidden and output layer, there are $1 \times 8 = 8$ weights and $1 \times 1 = 1$ bias. Altogether, there are 65 variables in this structure. In the classical model, these variables were adjusted using the Trainlm algorithm to train the MLPNN(6, 8, 1).

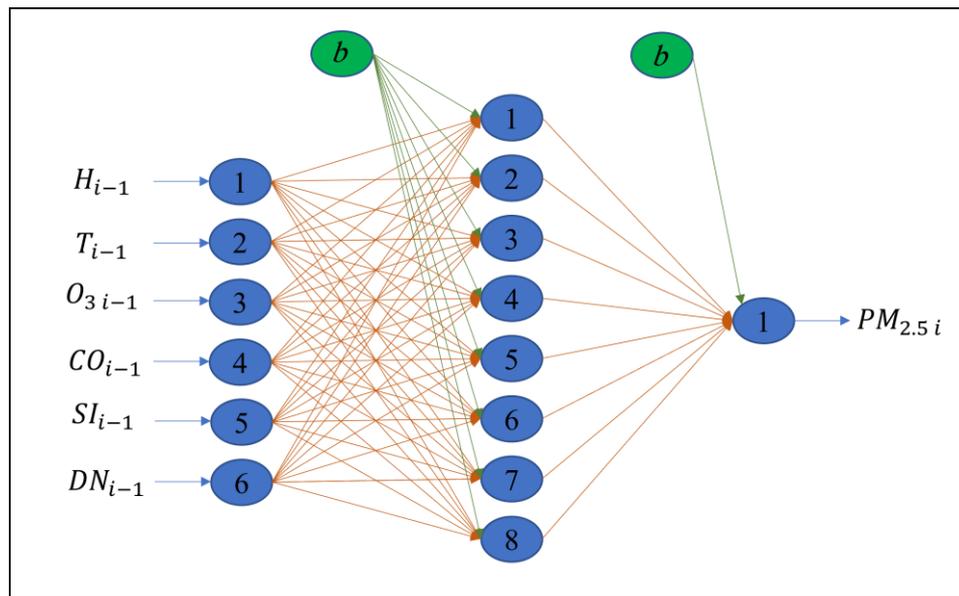


Figure 21: Topology of the used MLPNN(6, 8, 1) in Scenario 4

In the suggested hybrid model, the EFO replaces Trainlm to train the MLPNN(6, 8, 1). In other words, the EFO optimizes the weights and biases of the MLPNN(6, 8, 1) to preserve it from computational drawbacks. Hereafter, the created hybrid model is named EFO-MLPNN(6, 8, 1). Figure 22 shows the flowchart of the developed hybrid model. Generally speaking, optimization using a metaheuristic algorithm such as EFO requires taking the following steps:

- (i) Building the mathematical equation of the intended MLPNN(6, 8, 1)
- (ii) Introducing this equation as the problem function to the EFO
- (iii) Defining a cost function (here training RMSE to be minimized)
- (iv) Allocating the required parameters of the EFO including population size (S_{Pop}) and the number of iterations
- (v) Running the model to find the optimal solution within an iterative procedure

As per Figure 22, in a given iteration j , a new MLPNN(6, 8, 1) is created using the updated weights and biases (W.R.T iteration $j-1$). Its performance is then assessed by calculating $RMSE_j$. The stopping criterion is checked, and if it is not satisfied, iteration $j+1$ is deployed. This process continues until the algorithm reaches the maximum number of iterations [152].

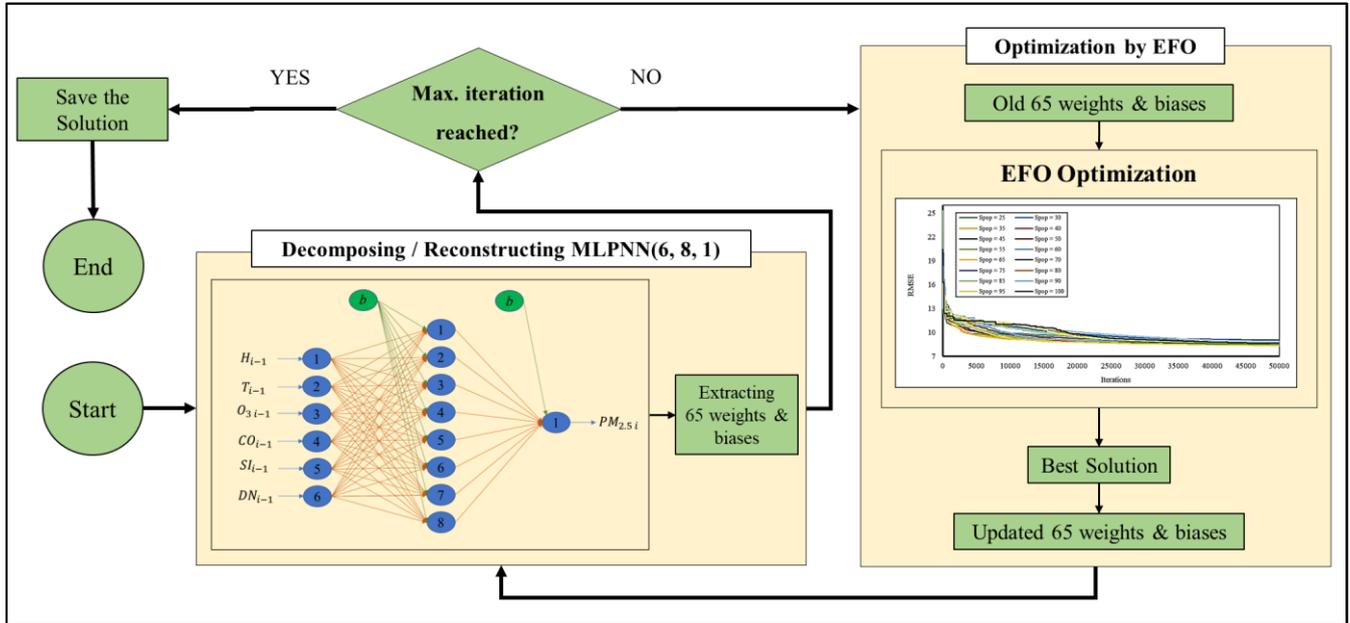


Figure 22: Hybridization of the MLPNN using EFO

Similar to the MLPNN, the EFO algorithm possesses some parameters that need to be assigned. A total of 50,000 iterations are determined after examining the behavior of this algorithm. Another parameter to be considered is S_{Pop} . It indicates the number of searching individuals in the problem space who seek the best solution (see Section 4.1.2.2). A trial-and-error process is performed to find the S_{Pop} that yields the most accurate results. The EFO-MLPNN(6, 8, 1) is run with sixteen candidate S_{Pop} s including 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, and 100 and their RMSEs are compared to identify the best option. Figure 23 shows the resultant optimization curves (collection of $RMSE_j$, $j = 1, 2, \dots, 50,000$). In addition, Figure 24 compares the obtained training and testing RMSEs for each S_{Pop} .

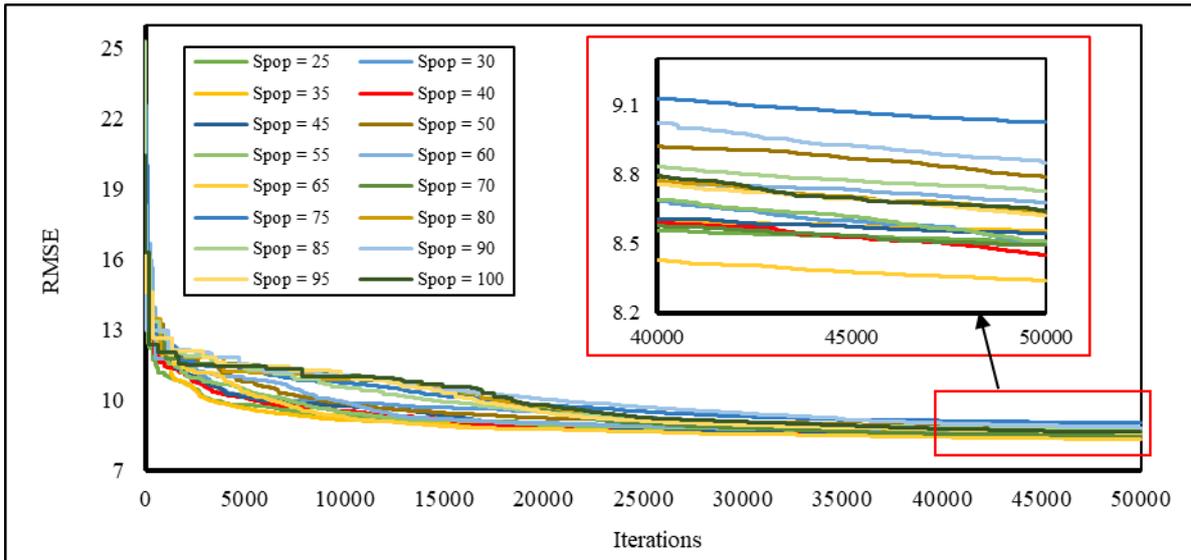


Figure 23: Optimization curves for the tried S_{Pop} s of the EFO algorithm

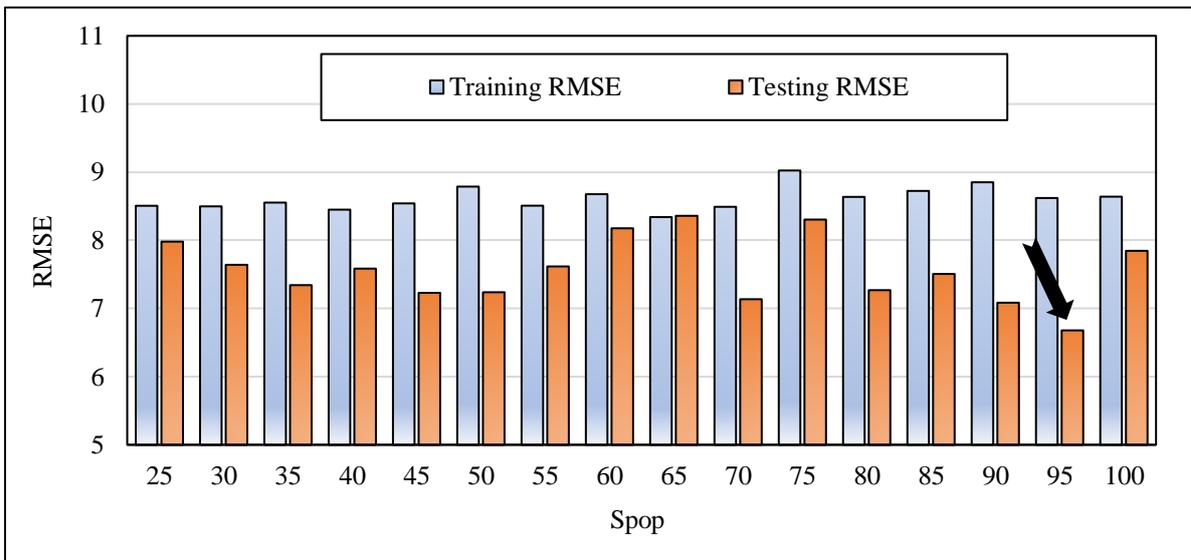


Figure 24: Sensitivity of training and testing RMSEs to the S_{Pop}

From Figure 23, all tried S_{Pop} s exhibit a compatible optimization path. The training error (i.e., RMSE) is reduced mostly in the first 20,000 iterations, and thereafter, the behavior of curves remains more or less stable. In the end, the curves reach a close RMSE. However, a comparison reveals the curve of $S_{Pop} = 65$ is slightly below the others (see the magnified section in Figure 23). It is also evident in Figure 24 (gray columns). While the level of training RMSEs is not very different throughout the chart, a significant distinction emerges when the testing RMSEs are

concerned. The smallest testing error is obtained for $S_{Pop} = 95$. Based on this reasoning, $S_{Pop} = 95$ is chosen, and therefore, the name of the selected model is updated as EFO(95)-MLPNN(6, 8, 1). Apart from the S_{Pop} and the number of iterations, the EFO has four user-assigned parameters, namely R_{rate} , $P_{S_{rate}}$, P_{field} , and N_{field} (see Section 4.1.2.2). These parameters were determined with reference to previous similar studies [153, 154], and their suitability was examined by trial implementations. Table 20 summarizes the EFO parameters.

Table 20: Parameters of the EFO algorithm

Parameter	Iterations	S_{Pop}	R_{rate}	$P_{S_{rate}}$	P_{field}	N_{field}
Value	50000	95	0.01	0.01	0.02	0.45

A notable advantage of the EFO compared to most metaheuristic algorithms is its fast calculations [155]. It enables the user to deploy a larger number of iterations to attain a more optimum solution. For this study, a 64-bit operating system with Intel (R) Core (TM) i78565U CPU @ 1.80GHz 1.99 GHz and 16 GB RAM is used. Implementing the EFO-MLPNN(6, 8, 1) algorithm in this system required by average 551 s (i.e., an average of the sixteen S_{Pop} s runs)

4.2.4 Prediction Results and Discussion

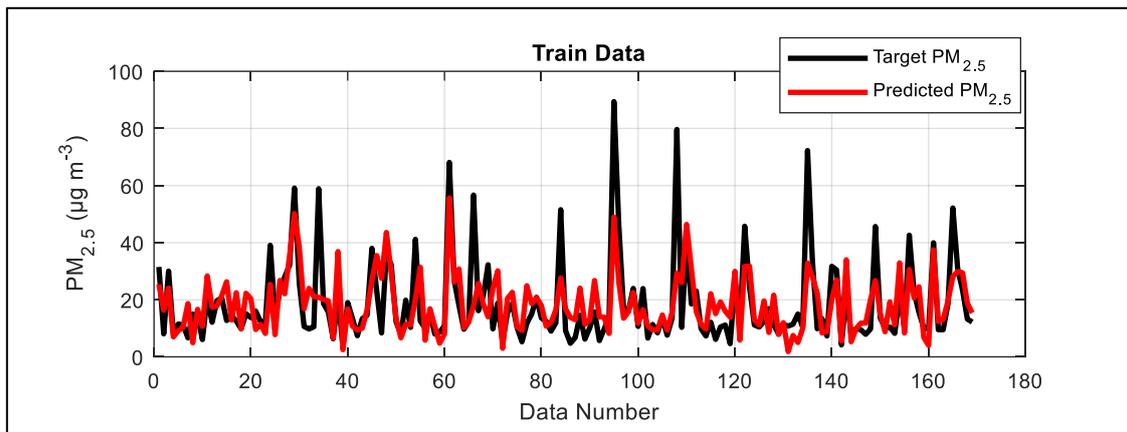
The prediction results are assessed in this section using the RMSE, MAE, and R_P criteria. The EFO(95)-MLPNN(6, 8, 1) is the pivotal model for predicting $PM_{2.5}$, while the MLPNN(6, 8, 1), as well as another ML technique called ANFIS, are considered as benchmarks.

The ANFIS was designed by Jang [156] and combines the training strategy of ANN and fuzzy theory. This model aims to map the input-output relationship in five steps dealing with (i) calculating input membership functions, (ii) calculating rule firing strength, (iii) normalizing rule firing strength, (iv) specifying node functions, and (v) calculating the final output [157]. Further mathematical details of this model can be found in earlier literature [158, 159]. The configuration of the used ANFIS depends on the number of clusters, which based on a trial-and-error process, is set to be 7 in this work. This model is therefore named ANFIS(7).

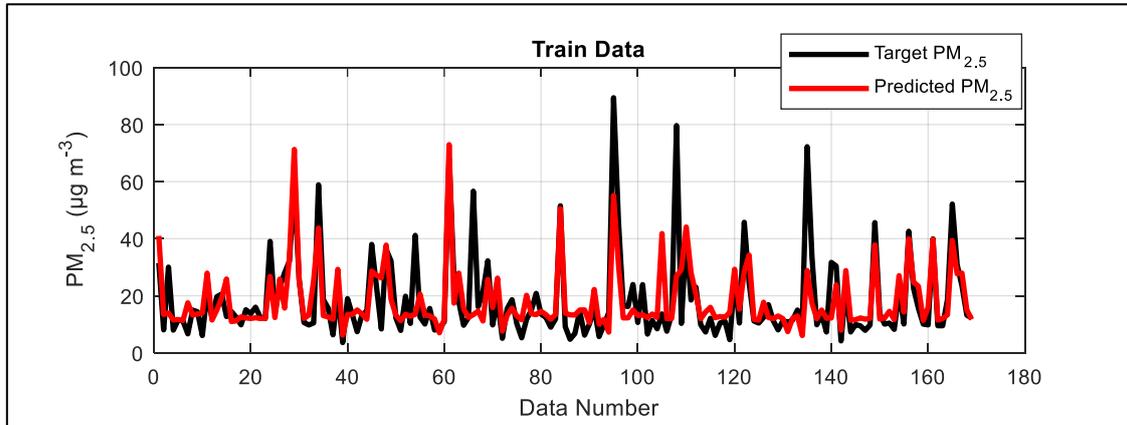
4.2.4.1 Training and Testing Accuracy

Figure 25 and Figure 26 present the training results. Figure 25 gives a comparison between the target and predicted trends of $PM_{2.5}$, and Figure 26 exhibits the corresponding histograms of error ($Error_i = PM_{2.5_{i_t}} - PM_{2.5_{i_o}}$) and correlation charts. The RMSEs 10.64, 9.77, and $8.62 \mu g m^{-3}$, along

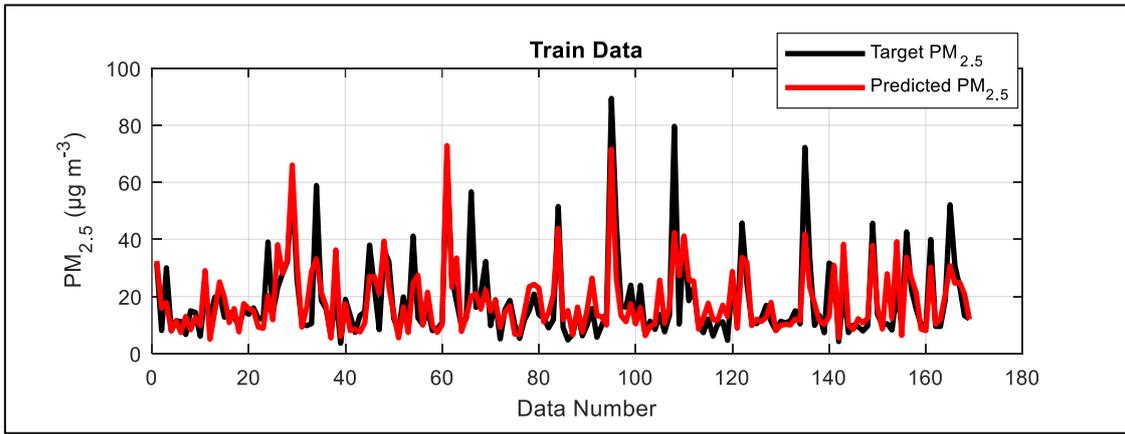
with the MAEs 7.10, 6.09, and 5.78 $\mu\text{g m}^{-3}$ are obtained for the ANFIS(7), MLPNN(6, 8, 1), and EFO(95)-MLPNN(6, 8, 1), respectively. These values indicate an acceptable level of accuracy for the training phase. The histogram charts also show a higher frequency of errors around minor magnitudes including zero. Referring to the calculated R_{PS} 0.69, 0.75, and 0.81, the correlation between the target and output $\text{PM}_{2.5}$ is effective. These results show that all three used models have learned the $\text{PM}_{2.5}$ behavior successfully. In other words, the dependency of $\text{PM}_{2.5}$ on the six input factors (i.e., H, T, O_3 , CO, SI, and DN) is nicely analyzed and understood by the ANFIS(7), MLPNN(6, 8, 1), and EFO(95)-MLPNN(6, 8, 1).



(a)

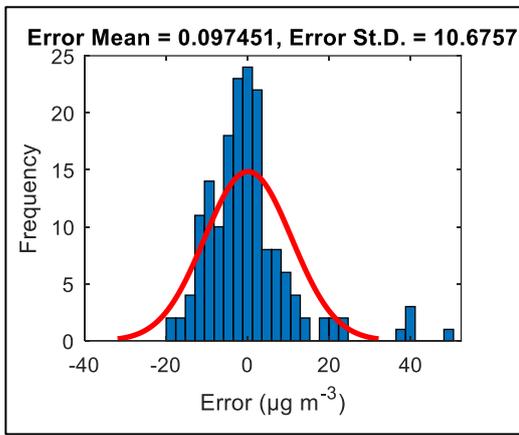


(b)

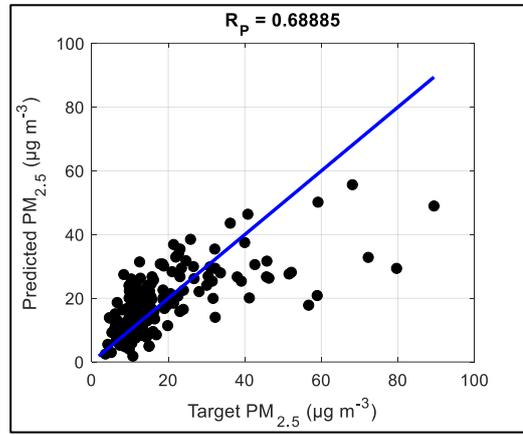


(c)

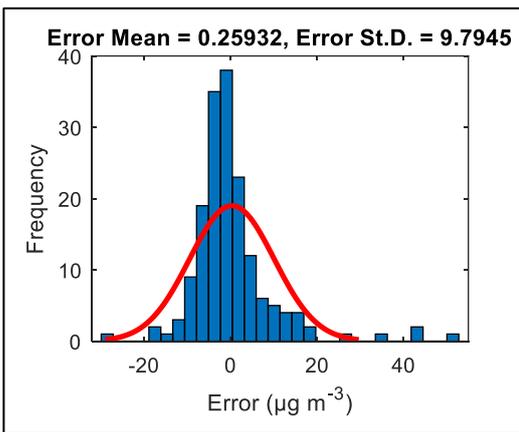
Figure 25: Target and predicted patterns of $PM_{2.5}$ in training phase of (a) ANFIS(7), (b) MLPNN(6, 8, 1), and (c) EFO(95)-MLPNN(6, 8, 1)



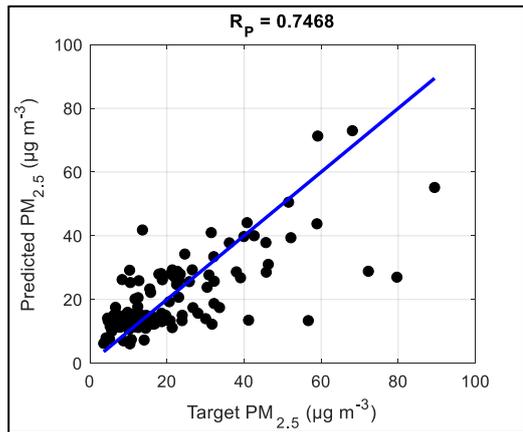
(a)



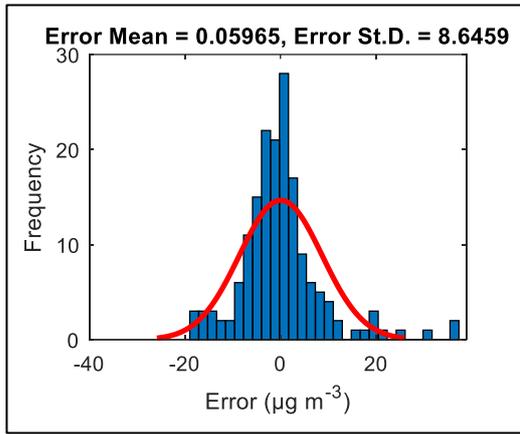
(b)



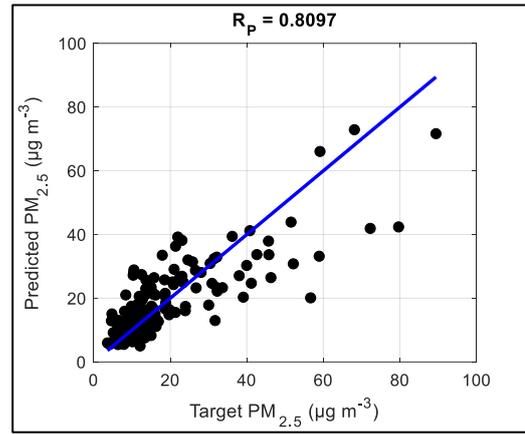
(c)



(d)



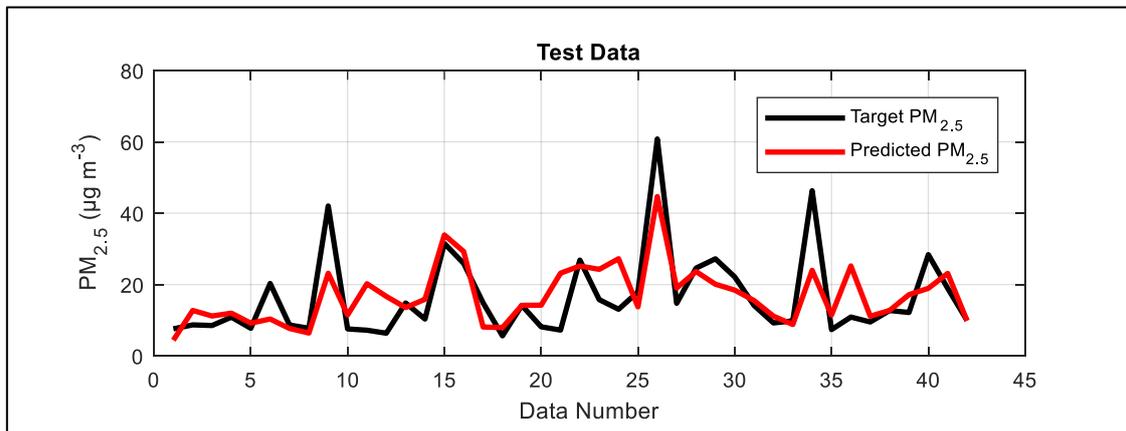
(e)



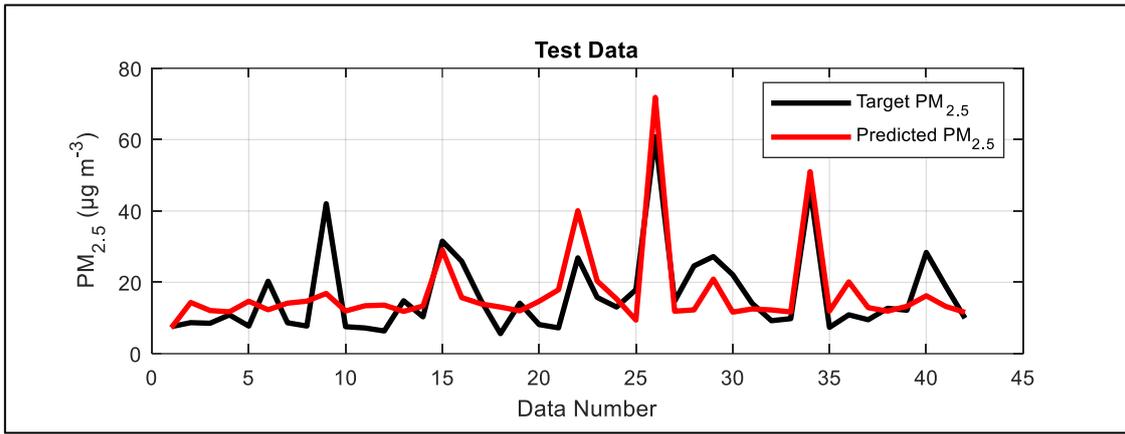
(f)

Figure 26: Histogram of error and correlation plots in the training phase of (a and b) ANFIS(7), (c and d) MLPNN(6, 8, 1), and (e and f) EFO(95)-MLPNN(6, 8, 1)

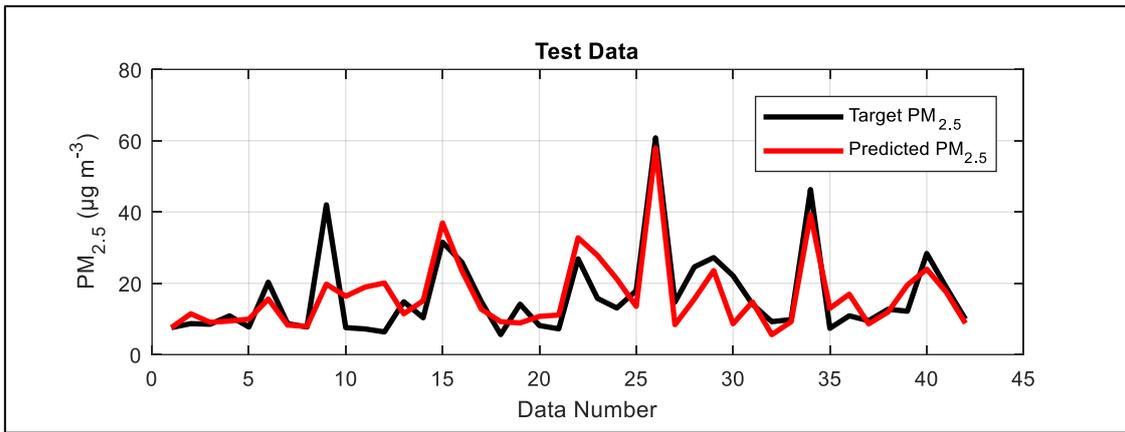
With the same representation, Figure 27 and Figure 28 illustrate the results of the testing phase. Based on the RMSEs 7.98, 7.51, and 6.68 $\mu\text{g m}^{-3}$, R_p s 0.73, 0.79, and 0.82, and also MAEs 5.72, 5.92, and 4.94 $\mu\text{g m}^{-3}$, it can be said that the prediction accuracy is satisfying. Graphically, Figure 27 shows that the general trend of $\text{PM}_{2.5}$ has been well followed by all predictive models. In addition, in Figure 28, histograms of error exhibit the aggregation of error frequencies around minor values including zero. The satisfying accordance between the target and output $\text{PM}_{2.5}$ s can be derived from the corresponding correlation charts.



(a)

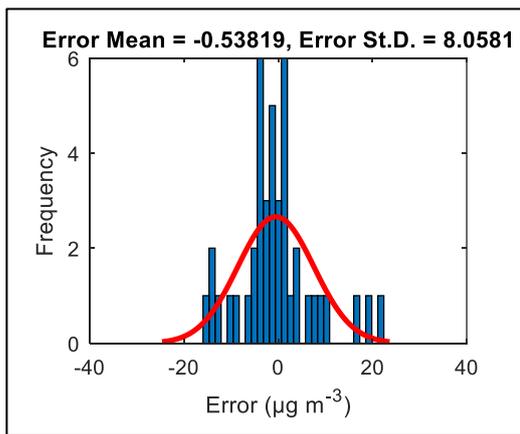


(b)

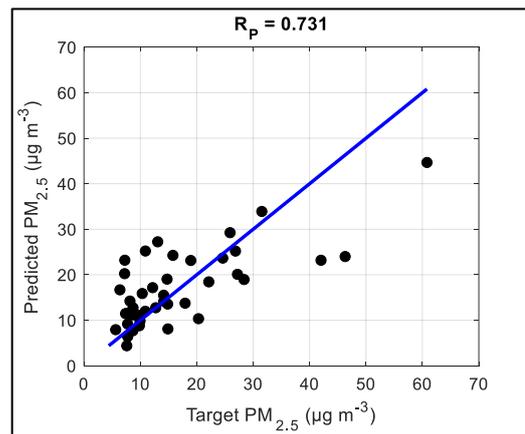


(c)

Figure 27: Target and predicted patterns of $PM_{2.5}$ in testing phase of (a) ANFIS(7), (b) MLPNN(6, 8, 1), and (c) EFO(95)-MLPNN(6, 8, 1)



(a)



(b)

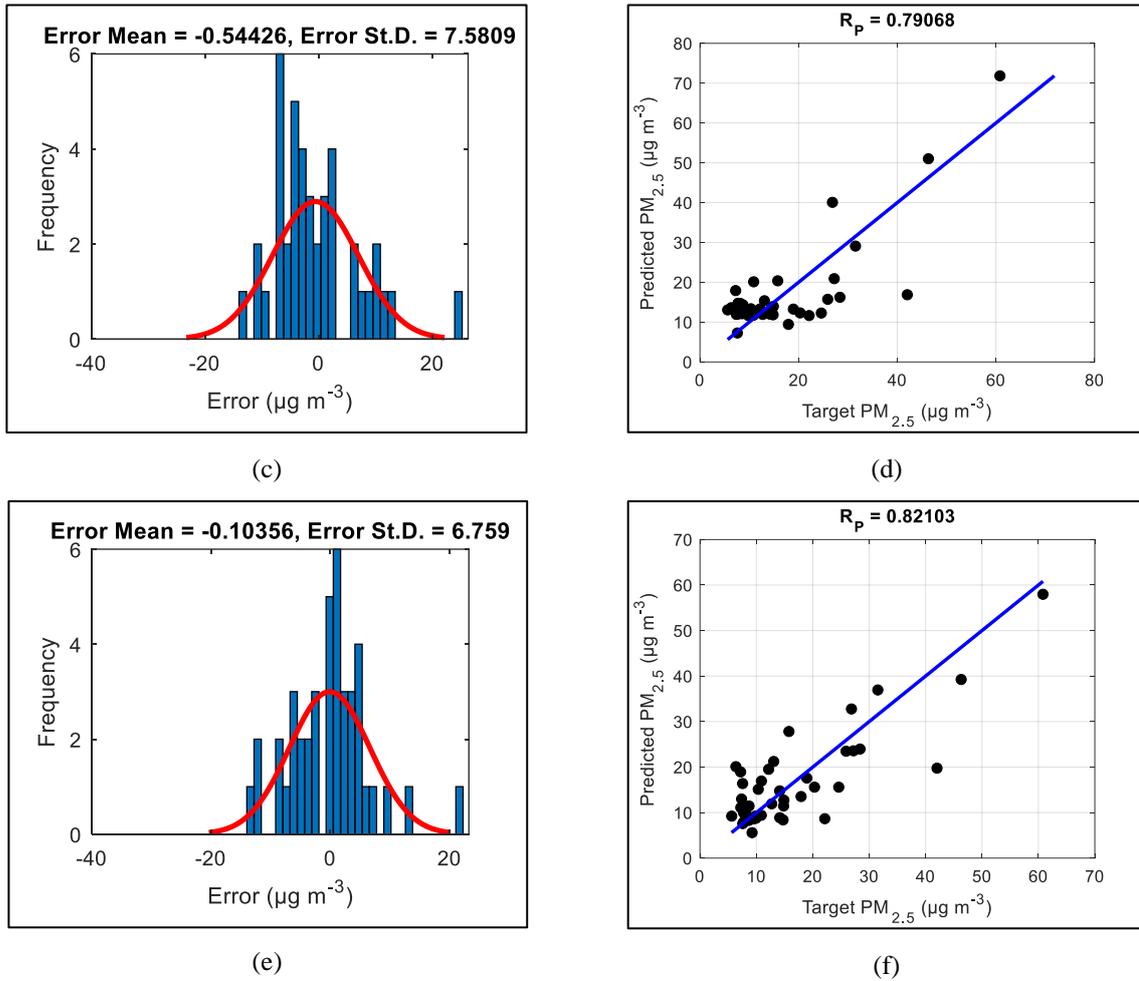


Figure 28: Histogram of error (a, c, e) and correlation plots (b, d, f) in the testing phase of (a and b) ANFIS(7), (c and d) MLPNN(6, 8, 1), and (e and f) EFO(95)-MLPNN(6, 8, 1)

4.2.4.2 Comparison

Table 21 gathers all calculated accuracy criteria for comparison purposes. Although all three models could achieve successful performance in both the training and testing phases, there are notable differences between their accuracy. At a glance, the EFO(95)-MLPNN(6, 8, 1) could achieve the most reliable analysis.

Table 21: Calculated accuracy criteria for training and testing phases

Model	Train			Test		
	RMSE	MAE	R _P	RMSE	MAE	R _P
ANFIS(7)	10.64	7.10	0.69	7.98	5.72	0.73
MLPNN(6, 8, 1)	9.77	6.09	0.75	7.51	5.92	0.79
EFO(95)-MLPNN(6, 8, 1)	8.62	5.78	0.81	6.68	4.94	0.82

In comparison with ANFIS(7), the training of EFO(95)-MLPNN(6, 8, 1) was around 18.98% and 18.59% more accurate in terms of the RMSE and MAE, respectively. Besides, its R_P indicates 12% higher correlation. As for testing results, EFO(95)-MLPNN(6, 8, 1) was around 16.29% and 13.63% more accurate in terms of the RMSE and MAE, respectively. Besides, its R_P indicates 9% higher correlation. In comparison with MLPNN(6, 8, 1), the training of EFO(95)-MLPNN(6, 8, 1) was around 11.77% and 5.09% more accurate in terms of the RMSE and MAE, respectively. Besides, its R_P indicates 6% higher correlation. As for testing results, EFO(95)-MLPNN(6, 8, 1) was around 11.05% and 16.55% more accurate in terms of the RMSE and MAE, respectively. Besides, its R_P indicates 3% higher correlation.

The above comparisons infer that optimizing the MLPNN(6, 8, 1) with the EFO algorithm has resulted in significant improvements. The better performance of the EFO(95)-MLPNN(6, 8, 1) in the training phase means that the EFO has helped ANN to enhance its understanding and reproduction of the $PM_{2.5}$ pattern. Similarly, the higher accuracy of the testing phase means that the hybrid model can better predict $PM_{2.5}$ for unseen meteorological, atmospheric, and temporal conditions. Furthermore, a comparative evaluation of the two benchmark models (i.e., MLPNN(6, 8, 1) vs. ANFIS(7)) ascertains the superiority of the MLPNN(6, 8, 1).

4.2.5 A Predictive Formula

Due to difficulties associated with implementing computer simulations, it is more convenient to use an explicit formula for predicting $PM_{2.5}$. In this section, a non-linear monolithic formula is extracted from the outstanding model of this study, i.e., EFO(95)-MLPNN(6, 8, 1). For doing this, 65 weights and biases of the MLPNN(6, 8, 1) that are optimized by EFO(95) are derived from the saved network (see Figure 22). Next, these weights and biases are organized with reference to Figure 21.

The single neuron in the output layer releases the $PM_{2.5}$ as expressed in Equation (9):

$$PM_{2.5} = 0.335119 \times LO_1 - 0.175616 \times LO_2 + 0.300288 \times LO_3 - 0.876373 \times LO_4 + 0.403931 \times LO_5 - 0.404279 \times LO_6 - 0.482643 \times LO_7 - 0.166935 \times LO_8 + 0.457510, \quad (9)$$

As Equation (9) denotes, this neuron first needs to receive local outputs (LO_i , $i = 1, 2, \dots, 8$). These eight values are the outcome of the eight neurons in the hidden layer. Equation (10) presents the calculations that are performed on the six input factors (i.e., H, T, O_3 , CO, SI, and DN) in the hidden layer for obtaining LO_i ($i = 1, 2, \dots, 8$).

$$\begin{matrix} LO_1 \\ LO_2 \\ LO_3 \\ LO_4 \\ LO_5 \\ LO_6 \\ LO_7 \\ LO_8 \end{matrix} = f \left(\begin{bmatrix} -0.263177 & 0.805952 & -1.101685 & -0.867334 & 0.468051 & 1.008033 \\ -1.300448 & -0.048378 & -0.918141 & -1.059940 & -0.473421 & -0.189503 \\ -0.102822 & 1.579138 & 0.145055 & 0.238365 & 1.146348 & 0.154197 \\ 0.350099 & 1.268840 & -0.001877 & -1.258985 & 0.047970 & -0.774679 \\ -0.472597 & 0.504687 & -0.738439 & -1.194986 & 1.176641 & 0.290129 \\ -0.148681 & 0.925231 & 0.966004 & 0.812428 & 0.388359 & -1.139217 \\ -0.919727 & -0.270874 & 1.115380 & -0.193985 & 0.245098 & -1.288003 \\ 0.145819 & 1.384931 & -0.284657 & -0.069436 & -0.061096 & 1.375180 \end{bmatrix} \times \begin{bmatrix} H \\ T \\ O_3 \\ CO \\ SI \\ DN \end{bmatrix} + \begin{bmatrix} 1.979899 \\ 1.414214 \\ 0.848528 \\ -0.282843 \\ -0.282843 \\ -0.848528 \\ -1.414214 \\ 1.979899 \end{bmatrix} \right), \quad (10)$$

where $f(x)$ is the Tansig activation function (see Table 14) that is defined by Equation (11):

$$f(x) = Tansig(x) = \frac{2}{1 + e^{-2x}} - 1, \quad (11)$$

To clarify the mechanism of this formula, it is identical to the process that takes place in the MLPNN model (see Figure 21). The MLPNN(6, 8, 1) first receives six inputs and performs the initial calculations in the hidden layer. As a result, eight neurons release eight LO_i ($i = 1, 2, \dots, 8$) and send them to the output layer. Finally, the output neuron calculates the $PM_{2.5}$. This formula can calculate $PM_{2.5}$ with the same accuracy as the EFO(95)-MLPNN(6, 8, 1). But there is a need to normalize the input data before using it. Consequently, the calculated $PM_{2.5}$ is not in the real range and must be denormalized accordingly [160].

4.2.6 Further Discussion

In Phase 2 of the study, a novel integrative methodology was proposed for the prediction of $PM_{2.5}$ as an air quality indicator. As per Section 4.2.2, when the MLPNN was executed in the classical form, in all cases, the training was stopped after a few epochs. In detail, while a total of 1000 epochs was permitted, the number of implemented epochs barely exceeded 20. The reason was the divergence of the solution which kept the model away from an optimum solution. This issue was resolved later by incorporating the EFO metaheuristic algorithm. In Figure 23, it is illustrated how the solution changes only if a better one is discovered (otherwise, the current solution remains). As a result, the accuracy of the hybrid model was considerably higher than the classical MLPNN.

Concerning the used optimizer, the EFO is one of the most efficient members of the metaheuristic family. This efficiency is interpreted as both great accuracy and fast implementation [153, 154]. In most – almost all – other metaheuristic algorithms, time is a critical parameter because optimization is carried out by large population sizes (e.g., $S_{Pop} > 100$ up to 1000) commonly within 1000 iterations that take a huge time. But the EFO, due to fast convergence and also utilizing lower S_{Pops}

allowed implementing 50 times more iterations within a reasonable time. However, conducting comparative studies for validating the EFO versus other existing metaheuristic algorithms would be of high interest for future works.

Despite that EFO(95)-MLPNN(6, 8, 1) could remedy some prediction weaknesses of its benchmark model, there are still some $PM_{2.5}$ underestimations that require attention. As it was depicted in Figure 25 and Figure 27, while EFO(95)-MLPNN(6, 8, 1) shows desirable sensitivity to the abrupt changes in the $PM_{2.5}$ behavior, a few peak values are underestimated (4 peaks in Figure 25 and 1 peak in Figure 27). For a more consistent illustration, Figure 29 collects the training and testing datasets to show the target and predicted $PM_{2.5}$ pattern in the whole period. According to this figure, the target $PM_{2.5}$ between May and October 2022 is less fluctuated and its prediction is in higher accordance compared to the initial and final parts of the chart. It denotes that the model may be susceptible to underestimating exceptionally high $PM_{2.5}$ values. In the future, this problem can be resolved by adding more samples of the peak values to the dataset which enables the model to broaden its knowledge.

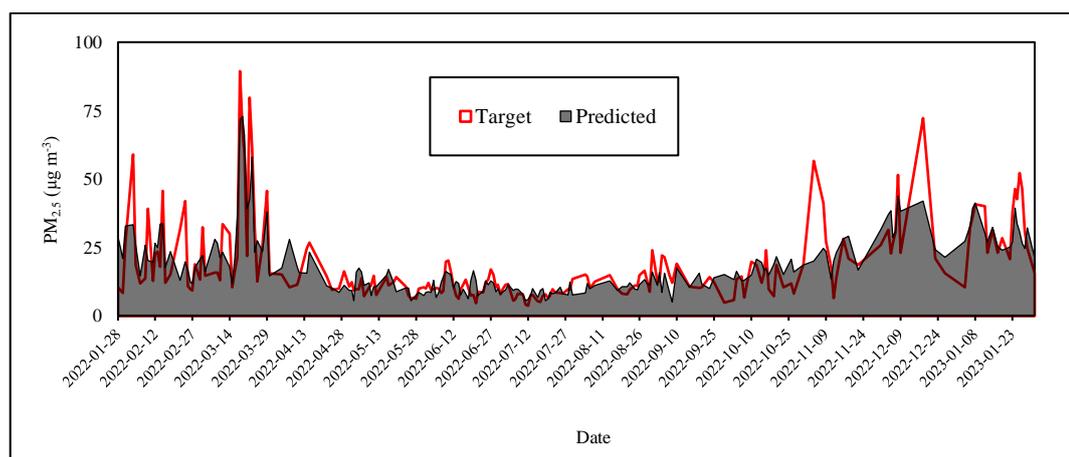


Figure 29: Comparing the target and predicted $PM_{2.5}$ trends during the studied period

Due to the limited ground station measurements in Ukraine, the second phase was developed only for Kiev. In addition, some parts of the used dataset were provided by Sentinel-5P measurements (i.e., atmospheric parameters including O_3 , NO_2 , and CO). As discussed in Phase 1, such data collection methods are associated with limitations such as insufficient observations and noised data. In fact, it was the reason that the effect of other gaseous pollutants (i.e., $HCHO$, CH_4 , and SO_2) on $PM_{2.5}$ was neglected in Phase 2. This drawback could reflect another potential idea for future works, i.e., evaluating how the prediction of air pollution is affected when further gaseous pollutants are taken into the equation.

Last but not least, in this work, $PM_{2.5}$ was a single representative of pollution. Taking other parameters into account may give a clearer understanding of air quality. A recommendation for this purpose is predicting the AQI which is a function of several parameters and exerting proper classification (e.g., from good to unhealthy conditions).

5. CONCLUSIONS

The Ukraine-Russia conflict has affected the surrounding environment from different aspects. This thesis was dedicated to investigating the air quality of Ukraine during this exceptional event. In the first phase, Sentinel-5P imagery was used to perform air quality monitoring (AQM). The variation in the emission of four gaseous pollutants (in 2022 vs. 2019 and 2021) including ozone (O_3), nitrogen dioxide (NO_2), formaldehyde (HCHO), and carbon monoxide (CO) in five major cities, namely Kiev, Kharkiv, Donetsk, Kherson, and Lviv was investigated. To sum up, evident changes were detected in the behavior of pollutants in all locations. These changes reflect the effect of war on anthropogenic activities in different parts of Ukraine. Moreover, comparing the results with similar previous literature showed suitable agreement between the findings. It is worth mentioning that this study could present some improvements concerning the involved parameters, studied cities, and the covered time span compared to the previous works.

In the second phase, machine learning models were developed to predict air pollution in Kiev. To this end, the concentration of $PM_{2.5}$ was predicted by a hybrid model composed of multi-layer perceptron neural network (MLPNN) as the basic predictor and electromagnetic field optimization (EFO) as the optimizer. In this process, $PM_{2.5}$ was a function of several meteorological, atmospheric, and temporal factors. A principal component analysis (PCA) was done to reduce the dimension of the problem. After evaluating four scenarios, sensitivity analyses pointed out the most effective configuration of the hybrid model. Based on the prediction results, the suggested EFO-MLPNN model could nicely comprehend and reproduce the relationship between the $PM_{2.5}$ and input factors. Moreover, a comparison showed that the hybrid model enjoys higher accuracy than the classical MLPNN, as well as adaptive neuro-fuzzy inference system (ANFIS). Comparison of the EFO-MLPNN vs. MLPNN demonstrated the excellent applicability of the EFO algorithm in optimizing the neural network for air quality prediction. An explicit formula was also derived from the EFO-MLPNN for a convenient prediction of $PM_{2.5}$.

In summary, this effort introduced the integration of Sentinel-5P imagery and artificial intelligence as a suitable approach for war-affected air quality assessment. However, the work encountered

some limitations, too, in terms of model development, and more importantly, data provision. Proper suggestions were accordingly given as clues for future projects to cope with these limitations.

6. BIBLIOGRAPHY

1. World_Bank, vol. no. pp. 2016.
2. M.S. Anjum, S.M. Ali, M.A. Subhani, M.N. Anwar, A.-S. Nizami et al., "An emerged challenge of air pollution and ever-increasing particulate matter in pakistan; a critical review," *Journal of Hazardous Materials*, vol. 402, no. pp. 123943, 2021.
3. S. Bočková, R. Bohovic, M. Hrnčiar, M. Muroň, P. Filippovová et al., "Air pollution in ukraine from space". Arnika–Citizen Support Centre, Prague, Czech Republic, 2020.
4. O. Popov, A. Iatsyshyn, V. Kovach, V. Artemchuk, I. Kameneva et al., "Risk assessment for the population of kyiv, ukraine as a result of atmospheric air pollution," *Journal of Health and Pollution*, vol. 10, no. 25, pp. 2020.
5. M. Savenets, "Air pollution in ukraine: A view from the sentinel-5p satellite," *Quarterly Journal of the Hungarian Meteorological Service*, vol. 125, no. 2, pp. 271-290, 2021.
6. G. Righini, A. Cappelletti, A. Ciucci, G. Cremona, A. Piersanti et al., "Gis based assessment of the spatial representativeness of air quality monitoring stations using pollutant emissions data," *Atmospheric Environment*, vol. 97, no. pp. 121-129, 2014.
7. R.V. Martin, "Satellite remote sensing of surface air quality," *Atmospheric Environment*, vol. 42, no. 34, pp. 7823-7843, 2008.
8. B.N. Duncan, A.I. Prados, L.N. Lamsal, Y. Liu, D.G. Streets et al., "Satellite data of atmospheric pollution for us air quality applications: Examples of applications, summary of data end-user resources, answers to faqs, and common mistakes to avoid," *Atmospheric Environment*, vol. 94, no. pp. 647-662, 2014.
9. S.S. Al-Alola, I.I. Alkadi, H.M. Alogayell, S.A. Mohamed and I.Y. Ismail, "Air quality estimation using remote sensing and gis-spatial technologies along al-shamal train pathway, al-qurayyat city in saudi arabia," *Environmental and Sustainability Indicators*, vol. no. pp. 100184, 2022.
10. C. Magro, L. Nunes, O.C. Gonçalves, N.R. Neng, J.M. Nogueira et al., "Atmospheric trends of co and ch4 from extreme wildfires in portugal using sentinel-5p tropomi level-2 data," *Fire*, vol. 4, no. 2, pp. 25, 2021.
11. B.W. Bodah, A. Neckel, L.S. Maculan, C.B. Milanes, C. Korcelski et al., "Sentinel-5p tropomi satellite application for no2 and co studies aiming at environmental valuation," *Journal of Cleaner Production*, vol. 357, no. pp. 131960, 2022.
12. L. Wang, M. Li, S. Yu, X. Chen, Z. Li et al., "Unexpected rise of ozone in urban and rural areas, and sulfur dioxide in rural areas during the coronavirus city lockdown in hangzhou, china: Implications for air quality," *Environmental Chemistry Letters*, vol. 18, no. 5, pp. 1713-1723, 2020.
13. A. Lorente, T. Borsdorff, A. Butz, O. Hasekamp, A. Schneider et al., "Methane retrieved from tropomi: Improvement of the data product and validation of the first 2 years of measurements," *Atmospheric Measurement Techniques*, vol. 14, no. 1, pp. 665-684, 2021.
14. T. Filippini, K.J. Rothman, A. Goffi, F. Ferrari, G. Maffei et al., "Satellite-detected tropospheric nitrogen dioxide and spread of sars-cov-2 infection in northern italy," *Science of The Total Environment*, vol. 739, no. pp. 140278, 2020.
15. V. Safarianzengir, B. Sobhani, M.H. Yazdani and M. Kianian, "Monitoring, analysis and spatial and temporal zoning of air pollution (carbon monoxide) using sentinel-5 satellite data for health management in iran, located in the middle east," *Air Quality, Atmosphere & Health*, vol. 13, no. 6, pp. 709-719, 2020.
16. C. Vigouroux, B. Langerock, C.A. Bauer Aquino, T. Blumenstock, Z. Cheng et al., "Tropomi–sentinel-5 precursor formaldehyde validation using an extensive network of ground-based fourier-transform infrared stations," *Atmospheric Measurement Techniques*, vol. 13, no. 7, pp. 3751-3767, 2020.

17. F. Zhao, C. Liu, Z. Cai, X. Liu, J. Bak et al., "Ozone profile retrievals from tropomi: Implication for the variation of tropospheric ozone during the outbreak of covid-19 in china," *Science of The Total Environment*, vol. 764, no. pp. 142886, 2021.
18. P. Broomandi, F. Karaca, A. Nikfal, A. Jahanbakhshi, M. Tamjidi et al., "Impact of covid-19 event on the air quality in iran," *Aerosol and Air Quality Research*, vol. 20, no. 8, pp. 1793-1804, 2020.
19. E.S.A. (ESA), vol. no. pp.
20. A.W. Ayoobi, H. Ahmadi, M. Inceoglu and E. Pekkan, "Seasonal impacts of buildings' energy consumption on the variation and spatial distribution of air pollutant over kabul city: Application of sentinel—5p tropomi products," *Air Quality, Atmosphere & Health*, vol. 15, no. 1, pp. 73-83, 2022.
21. R. Bhatkar, S.R.N.A. Syamala and J.T. Varghese. "Impact of population count on the presence of nitrogen dioxide in united arab emirates using sentinel-5p satellite data". in *2020 IEEE 15th International Conference on Industrial and Information Systems (ICIIS)*. 2020. IEEE.
22. L. Guanter, C. Bacour, A. Schneider, I. Aben, T.A. van Kempen et al., "The tropisif global sun-induced fluorescence dataset from the sentinel-5p tropomi mission," *Earth System Science Data*, vol. 13, no. 11, pp. 5423-5440, 2021.
23. T. K uchler, S. No el, H. Bovensmann, J.P. Burrows, T. Wagner et al., "Total water vapour columns derived from sentinel 5p using the amc-doas method," *Atmospheric Measurement Techniques*, vol. 15, no. 2, pp. 297-320, 2022.
24. C. Liu, M. Gao, Q. Hu, G.P. Brasseur and G.R. Carmichael, "Stereoscopic monitoring: A promising strategy to advance diagnostic and prediction of air pollution," *Bulletin of the American Meteorological Society*, vol. 102, no. 4, pp. E730-E737, 2021.
25. O.S. Yilmaz, U. Acar, F.B. Sanli, F. Gulgen and A.M. Ates, "Mapping burn severity and monitoring co content in t urkiye's 2021 wildfires, using sentinel-2 and sentinel-5p satellite data on the gee platform," *Earth Science Informatics*, vol. no. pp. 1-20, 2023.
26. A. Cofano, F. Cigna, L. Santamaria Amato, M. Siciliani de Cumis and D. Tapete, "Exploiting sentinel-5p tropomi and ground sensor data for the detection of volcanic so2 plumes and activity in 2018–2021 at stromboli, italy," *Sensors*, vol. 21, no. 21, pp. 6991, 2021.
27. D. De Santis, I. Petracca, S. Corradini, L. Guerrieri, M. Picchiani et al. "Volcanic so2 near-real time retrieval using tropomi data and neural networks: The december 2018 etna test case". in *2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS*. 2021. IEEE.
28. A. Sharifi and S. Felegari, "Nitrogen dioxide (no2) pollution monitoring with sentinel-5p satellite imagery over during the coronavirus pandemic (case study: Tehran)," *Remote Sensing Letters*, vol. 13, no. 10, pp. 1029-1039, 2022.
29. M. Virghileanu, I. S avulescu, B.-A. Mihai, C. Nistor and R. Dobre, "Nitrogen dioxide (no2) pollution monitoring with sentinel-5p satellite imagery over europe during the coronavirus pandemic outbreak," *Remote Sensing*, vol. 12, no. 21, pp. 3575, 2020.
30. K. Muniraj, B. Panneerselvam, S. Devaraj, C.J. Jesudhas and K. Sudalaimuthu, "Evaluating the effectiveness of emissions reduction measures and ambient air quality variability through ground-based and sentinel-5p observations under the auspices of covid pandemic lockdown in tamil nadu, india," *International Journal of Environmental Analytical Chemistry*, vol. no. pp. 1-12, 2021.
31. D. Stratoulas and N. Nuthammachot, "Air quality development during the covid-19 pandemic over a medium-sized urban area in thailand," *Science of The Total Environment*, vol. 746, no. pp. 141320, 2020.
32. P.F. Levelt, D.C. Stein Zweers, I. Aben, M. Bauwens, T. Borsdorff et al., "Air quality impacts of covid-19 lockdown measures detected from space using high spatial resolution observations of multiple trace gases from sentinel-5p/tropomi," *Atmospheric Chemistry and Physics*, vol. 22, no. 15, pp. 10319-10351, 2022.

33. M.I. Volke, R. Abarca-del-Rio and C. Ulloa-Tesser, "Impact of mobility restrictions on no₂ concentrations in key latin american cities during the first wave of the covid-19 pandemic," *Urban Climate*, vol. no. pp. 101412, 2023.
34. S.P. Mandal, A.K. Ranjan, B.R. Parida and S.N. Behera, "Assessing aerosol and nitrogen dioxide concentration in major urban cities over the himalayan region during the covid-19 lockdown phases", in *Handbook of himalayan ecosystems and sustainability, volume 2*. 2023, CRC Press. p. 293-315.
35. T. Husain, "Terrestrial and atmospheric environment during and after the gulf war," *Environment international*, vol. 24, no. 1-2, pp. 189-196, 1998.
36. P. Pereira, F. Bašić, I. Bogunovic and D. Barcelo, "Russian-ukrainian war impacts the total environment," *Science of The Total Environment*, vol. 837, no. pp. 155865, 2022.
37. T. Ali, M. Abouleish, R. Gawai, N. Hamdan and A. Elaksher, "Ammonium nitrate explosion at the main port in beirut (lebanon) and air pollution: An analysis of the spatiotemporal distribution of nitrogen dioxide," *Euro-Mediterranean Journal for Environmental Integration*, vol. 7, no. 1, pp. 21-27, 2022.
38. O.Y. Al-Jarrah, P.D. Yoo, S. Muhaidat, G.K. Karagiannidis and K. Taha, "Efficient machine learning for big data: A review," *Big Data Research*, vol. 2, no. 3, pp. 87-93, 2015.
39. J.G. Carbonell, R.S. Michalski and T.M. Mitchell, "An overview of machine learning," *Machine learning*, vol. no. pp. 3-23, 1983.
40. M. Flah, I. Nunez, W. Ben Chaabene and M.L. Nehdi, "Machine learning algorithms in civil structural health monitoring: A systematic review," *Archives of computational methods in engineering*, vol. 28, no. pp. 2621-2643, 2021.
41. P. Tahmasebi, S. Kamrava, T. Bai and M. Sahimi, "Machine learning in geo-and environmental sciences: From small to large scale," *Advances in Water Resources*, vol. 142, no. pp. 103619, 2020.
42. L. Zheng, R. Lin, X. Wang and W. Chen, "The development and application of machine learning in atmospheric environment studies," *Remote Sensing*, vol. 13, no. 23, pp. 4839, 2021.
43. C. Bellinger, M.S. Mohamed Jabbar, O. Zaïane and A. Osornio-Vargas, "A systematic review of data mining and machine learning for air pollution epidemiology," *BMC public health*, vol. 17, no. pp. 1-19, 2017.
44. Y.A. Ayturan, Z.C. Ayturan and H.O. Altun, "Air pollution modelling with deep learning: A review," *International Journal of Environmental Pollution and Environmental Modelling*, vol. 1, no. 3, pp. 58-62, 2018.
45. Y. Rybarczyk and R. Zalakeviciute, "Machine learning approaches for outdoor air quality modelling: A systematic review," *Applied Sciences*, vol. 8, no. 12, pp. 2570, 2018.
46. D. Iskandaryan, F. Ramos and S. Trilles, "Air quality prediction in smart cities using machine learning technologies based on sensor data: A review," *Applied Sciences*, vol. 10, no. 7, pp. 2401, 2020.
47. S.M. Cabaneros, J.K. Calautit and B.R. Hughes, "A review of artificial neural network models for ambient air pollution prediction," *Environmental Modelling & Software*, vol. 119, no. pp. 285-304, 2019.
48. M. Zeinalnezhad, A.G. Chofreh, F.A. Goni and J.J. Klemeš, "Air pollution prediction using semi-experimental regression model and adaptive neuro-fuzzy inference system," *Journal of Cleaner Production*, vol. 261, no. pp. 121218, 2020.
49. W. Lu, W. Wang, A.Y. Leung, S.-M. Lo, R.K. Yuen et al. "Air pollutant parameter forecasting using support vector machines". in *Proceedings of the 2002 International Joint Conference on Neural Networks. IJCNN'02 (Cat. No. 02CH37290)*. 2002. IEEE.
50. R. Feng, H.-j. Zheng, H. Gao, A.-r. Zhang, C. Huang et al., "Recurrent neural network and random forest for analysis and accurate forecast of atmospheric pollutants: A case study in hangzhou, china," *Journal of Cleaner Production*, vol. 231, no. pp. 1005-1015, 2019.

51. H. Maleki, A. Sorooshian, G. Goudarzi, Z. Baboli, Y. Tahmasebi Birgani et al., "Air pollution prediction by using an artificial neural network model," *Clean technologies and environmental policy*, vol. 21, no. pp. 1341-1352, 2019.
52. Y. Zhang, Y. Wang, M. Gao, Q. Ma, J. Zhao et al., "A predictive data feature exploration-based air quality prediction approach," *IEEE Access*, vol. 7, no. pp. 30732-30743, 2019.
53. N.a. Zaini, L.W. Ean, A.N. Ahmed and M.A. Malek, "A systematic literature review of deep learning neural network for time series air quality forecasting," *Environmental Science and Pollution Research*, vol. no. pp. 1-33, 2022.
54. M. Kaveh and M.S. Mesgari, "Application of meta-heuristic algorithms for training neural networks and deep learning architectures: A comprehensive review," *Neural Processing Letters*, vol. no. pp. 1-104, 2022.
55. H.R. Maier, S. Razavi, Z. Kapelan, L.S. Matott, J. Kasprzyk et al., "Introductory overview: Optimization using evolutionary algorithms and other metaheuristics," *Environmental Modelling & Software*, vol. 114, no. pp. 195-213, 2019.
56. H. Liu, G. Yan, Z. Duan and C. Chen, "Intelligent modeling strategies for forecasting air quality time series: A review," *Applied Soft Computing*, vol. 102, no. pp. 106957, 2021.
57. G.I. Drewil and R.J. Al-Bahadili, "Air pollution prediction using lstm deep learning and metaheuristics algorithms," *Measurement: Sensors*, vol. 24, no. pp. 100546, 2022.
58. A. Yonar and H. Yonar, "Modeling air pollution by integrating anfis and metaheuristic algorithms," *Modeling Earth Systems and Environment*, vol. no. pp. 1-11, 2022.
59. A. Heydari, M. Majidi Nezhad, D. Astiaso Garcia, F. Keynia and L. De Santoli, "Air pollution forecasting application based on deep learning model and optimization algorithm," *Clean technologies and environmental policy*, vol. no. pp. 1-15, 2021.
60. S. Zhu, X. Qiu, Y. Yin, M. Fang, X. Liu et al., "Two-step-hybrid model based on data preprocessing and intelligent optimization algorithms (cs and gwo) for no2 and so2 forecasting," *Atmospheric Pollution Research*, vol. 10, no. 4, pp. 1326-1335, 2019.
61. Z. Kang and Z. Qu. "Application of bp neural network optimized by genetic simulated annealing algorithm to prediction of air quality index in lanzhou". in *2017 2nd IEEE international conference on computational intelligence and applications (ICCI)*. 2017. IEEE.
62. X. Xia. "Study on the application of bp neural network in air quality prediction based on adaptive chaos fruit fly optimization algorithm". in *MATEC Web of Conferences*. 2021. EDP Sciences.
63. O. Ovdiienko, M. Hryhorak, V. Marchuk and D. Bugayko, "An assessment of the aviation industry's impact on air pollution from its emissions: Worldwide and the ukraine," *Environmental & Socio-economic Studies*, vol. 9, no. 2, pp. 1-10, 2021.
64. L. Menut, B. Bessagnet, G. Siour, S. Mailler, R. Pennel et al., "Impact of lockdown measures to combat covid-19 on air quality over western europe," *Science of The Total Environment*, vol. 741, no. pp. 140426, 2020.
65. D. Rawtani, G. Gupta, N. Khatri, P.K. Rao and C.M. Hussain, "Environmental damages due to war in ukraine: A perspective," *Science of The Total Environment*, vol. 850, no. pp. 157932, 2022.
66. C. Zhang, Q. Hu, W. Su, C. Xing and C. Liu, "Satellite spectroscopy reveals the atmospheric consequences of the 2022 russia-ukraine war," *Science of The Total Environment*, vol. no. pp. 161759, 2023.
67. R. Zalakeviciute, D. Mejia, H. Alvarez, X. Bermeo, S. Bonilla-Bedoya et al., "War impact on air quality in ukraine," *Sustainability*, vol. 14, no. 21, pp. 13832, 2022.
68. T.W.B. Climatology, vol. no. pp.
69. A. Shvidenko, I. Buksha, S. Krakovska and P. Lakyda, "Vulnerability of ukrainian forests to climate change," *Sustainability*, vol. 9, no. 7, pp. 1152, 2017.
70. T.W.B. Population, "

" vol. no. pp.

71. G.B. Piccoli, G. Brunori, L. Gesualdo and K. Kalantar-Zadeh, "The impact of the russian–ukrainian war for people with chronic diseases," *Nature Reviews Nephrology*, vol. 18, no. 7, pp. 411-412, 2022.
72. Reuters, "Timeline: The events leading up to russia's invasion of ukraine," vol. no. pp.
73. Statista, "Number of civilian casualties in ukraine during russia's invasion verified by ohchr as of january 29, 2023," vol. no. pp.
74. BBC, "Ukraine war: Us estimates 200,000 military casualties on all sides
" vol. no. pp.
75. Euronews, "Russia’s war will shrink ukraine’s economy by 45% this year, world bank says," vol. no. pp.
76. I.f.t.S.o.W.a.A.s.C.T. Project, vol. no. pp.
77. Humdata.org, "Ukraine data explorer, total number of conflict events," vol. no. pp.
78. NPR, "U.S., nato countries announce massive weapons package for ukraine," vol. no. pp.
79. Euronews, "Ukraine war: Which weapons are being used in russia's invasion?," vol. no. pp.
80. Wikipedia, "List of russo-ukrainian conflict military equipment," vol. no. pp.
81. Washingtonpost, "The russian weaponry being used to attack ukraine," vol. no. pp.
82. M.A. Ghannadi, M. Shahri, S. Alebooyeh and A. Moradi, "Evaluation of sulfur dioxide emissions in thermal power plant and its effect on air quality in the neighboring city using sentinel-5 images (case study: Iran, arak)," *Earth Observation and Geomatics Engineering*, vol. 5, no. 1, pp. 36-45, 2021.
83. L. Shikwambana, P. Mhangara and N. Mbatha, "Trend analysis and first time observations of sulphur dioxide and nitrogen dioxide in south africa using tropomi/sentinel-5 p data," *International Journal of Applied Earth Observation and Geoinformation*, vol. 91, no. pp. 102130, 2020.
84. E.E.S. Agency), "Sentinel-5p," vol. no. pp.
85. C.O.A. Hub, vol. no. pp.
86. S.-P.P.-O.D. Hub, vol. no. pp.
87. E.B. Sentinel Hub, vol. no. pp.
88. T.E. Earth Engine Apps, vol. no. pp.
89. M. Amani, A. Ghorbanian, S.A. Ahmadi, M. Kakooei, A. Moghimi et al., "Google earth engine cloud computing platform for remote sensing big data applications: A comprehensive review," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 13, no. pp. 5326-5350, 2020.
90. Q. Zhao, L. Yu, X. Li, D. Peng, Y. Zhang et al., "Progress and trends in the application of google earth and google earth engine," *Remote Sensing*, vol. 13, no. 18, pp. 3778, 2021.
91. H. Tamiminia, B. Salehi, M. Mahdianpari, L. Quackenbush, S. Adeli et al., "Google earth engine for geo-big data applications: A meta-analysis and systematic review," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 164, no. pp. 152-170, 2020.
92. B. Hao, M. Ma, S. Li, Q. Li, D. Hao et al., "Land use change and climate variation in the three gorges reservoir catchment from 2000 to 2015 based on the google earth engine," *Sensors*, vol. 19, no. 9, pp. 2118, 2019.
93. E.L. Bullock, C.E. Woodcock and P. Olofsson, "Monitoring tropical forest degradation using spectral unmixing and landsat time series analysis," *Remote Sensing of Environment*, vol. 238, no. pp. 110968, 2020.
94. P. Gong, X. Li, J. Wang, Y. Bai, B. Chen et al., "Annual maps of global artificial impervious area (gaia) between 1985 and 2018," *Remote Sensing of Environment*, vol. 236, no. pp. 111510, 2020.

95. H. Huang, Y. Chen, N. Clinton, J. Wang, X. Wang et al., "Mapping major land cover dynamics in beijing using all landsat images in google earth engine," *Remote Sensing of Environment*, vol. 202, no. pp. 166-176, 2017.
96. D.K. Matci, G. Kaplan and U. Avdan, "Changes in air quality over different land covers associated with covid-19 in turkey aided by gee," *Environmental monitoring and assessment*, vol. 194, no. 10, pp. 762, 2022.
97. L. Kumar and O. Mutanga, "Google earth engine applications since inception: Usage, trends, and potential," *Remote Sensing*, vol. 10, no. 10, pp. 1509, 2018.
98. L. Curtis, W. Rea, P. Smith-Willis, E. Fenyves and Y. Pan, "Adverse health effects of outdoor air pollutants," *Environment international*, vol. 32, no. 6, pp. 815-830, 2006.
99. D. Zhu, C. Cai, T. Yang and X. Zhou, "A machine learning approach for air quality prediction: Model regularization and optimization," *Big data and cognitive computing*, vol. 2, no. 1, pp. 5, 2018.
100. E.E.D. Catalog-Sentinel-5P, vol. no. pp.
101. E.S.A. (ESA), "Mission status reports," vol. no. pp.
102. M.P.C. Tropomi, vol. no. pp.
103. N.M. Napi, S. Abdullah, A. Ahmed, A.A. Mansor and M. Ismail. "Annual and diurnal trend of surface ozone (o₃) in industrial area". in *IOP Conference Series: Earth and Environmental Science*. 2020. IOP Publishing.
104. J.J. West, A.M. Fiore, L.W. Horowitz and D.L. Mauzerall, "Global health benefits of mitigating ozone pollution with methane emission controls," *Proceedings of the National Academy of Sciences*, vol. 103, no. 11, pp. 3988-3993, 2006.
105. D. Nuvolone, D. Petri and F. Voller, "The effects of ozone on human health," *Environmental Science and Pollution Research*, vol. 25, no. pp. 8074-8088, 2018.
106. E.E. Agency, "Air quality in europe — 2018 report," vol. no. pp. 2018.
107. G.D. Nielsen and P. Wolkoff, "Cancer effects of formaldehyde: A proposal for an indoor air guideline value," *Archives of toxicology*, vol. 84, no. pp. 423-446, 2010.
108. I.A.f.R.o. Cancer, "Formaldehyde, 2-butoxyethanol and 1-tert-butoxypropan-2-ol", in *Formaldehyde, 2-butoxyethanol and 1-tert-butoxypropan-2-ol*. 2006. p. 478-478.
109. R.C. Gupta, "Handbook of toxicology of chemical warfare agents," in ed., Place, Academic Press, 2015.
110. J. Integrated Science Assessment (ISA) for Carbon Monoxide (Final Report, vol. no. pp.
111. E.E.S. Agency), "Sentinel-5p pre-operations data hub - european.," vol. no. pp. 2020.
112. E.E.D.C.-. Sentinel-5P, vol. no. pp.
113. R. Spurr, M. Van Roozendaal, C. Fayt and J.-C. Lambert. "*The godfit (gome direct fitting) algorithm: A new approach for total column retrieval*". in *Envisat & ERS Symposium*. 2005.
114. T. Explorer, vol. no. pp.
115. R.S. Sokhi, V. Singh, X. Querol, S. Finardi, A.C. Targino et al., "A global observational analysis to understand changes in air quality during exceptionally low anthropogenic emission conditions," *Environment international*, vol. 157, no. pp. 106818, 2021.
116. D. Griffin, X. Zhao, C.A. McLinden, F. Boersma, A. Bourassa et al., "High-resolution mapping of nitrogen dioxide with tropomi: First results and validation over the canadian oil sands," *Geophysical Research Letters*, vol. 46, no. 2, pp. 1049-1060, 2019.
117. G. Kaplan and Z.Y. Avdan, "Space-borne air pollution observation from sentinel-5p tropomi: Relationship between pollutants, geographical and demographic data," *International Journal of Engineering and Geosciences*, vol. 5, no. 3, pp. 130-137, 2020.
118. D.R. Leasure, R. Kashyap, F. Rampazzo, B. Elbers, C. Dooley et al., "Ukraine crisis: Monitoring population displacement through social media activity," vol. no. pp. 2022.

119. M.C. Collivignarelli, A. Abbà, G. Bertanza, R. Pedrazzani, P. Ricciardi et al., "Lockdown for covid-2019 in milan: What are the effects on air quality?," *Science of The Total Environment*, vol. 732, no. pp. 139280, 2020.
120. L.Y.K. Nakada and R.C. Urban, "Covid-19 pandemic: Impacts on the air quality during the partial lockdown in são paulo state, brazil," *Science of The Total Environment*, vol. 730, no. pp. 139087, 2020.
121. M.G. Adam, P.T. Tran and R. Balasubramanian, "Air quality changes in cities during the covid-19 lockdown: A critical review," *Atmospheric Research*, vol. 264, no. pp. 105823, 2021.
122. M. Mehrabi and H. Moayedi, "Landslide susceptibility mapping using artificial neural network tuned by metaheuristic algorithms," *Environmental Earth Sciences*, vol. 80, no. 24, pp. 1-20, 2021.
123. G.M. Marcazzan, S. Vaccaro, G. Valli and R. Vecchi, "Characterisation of pm10 and pm2. 5 particulate matter in the ambient air of milan (italy)," *Atmospheric Environment*, vol. 35, no. 27, pp. 4639-4650, 2001.
124. N.R. Martins and G.C. Da Graca, "Impact of pm2. 5 in indoor urban environments: A review," *Sustainable Cities and Society*, vol. 42, no. pp. 259-275, 2018.
125. Q. Liao, M. Zhu, L. Wu, X. Pan, X. Tang et al., "Deep learning for air quality forecasts: A review," *Current Pollution Reports*, vol. 6, no. pp. 399-409, 2020.
126. PurpleAir, "Shepeleva sensor," vol. no. pp.
127. Solcast, vol. no. pp.
128. K.K. Barkjohn, B. Gantt and A.L. Clements, "Development and application of a united states-wide correction for pm 2.5 data collected with the purpleair sensor," *Atmospheric Measurement Techniques*, vol. 14, no. 6, pp. 4617-4637, 2021.
129. O. Seyedashraf, M. Mehrabi and A.A. Akhtari, "Novel approach for dam break flow modeling using computational intelligence," *Journal of Hydrology*, vol. 559, no. pp. 1028-1038, 2018.
130. H. Moayedi, M. Mehrabi, D.T. Bui, B. Pradhan and L.K. Foong, "Fuzzy-metaheuristic ensembles for spatial assessment of forest fire susceptibility," *Journal of environmental management*, vol. 260, no. pp. 109867, 2020.
131. W.S. McCulloch and W. Pitts, "A logical calculus of the ideas immanent in nervous activity," *The bulletin of mathematical biophysics*, vol. 5, no. pp. 115-133, 1943.
132. K. Hornik, "Approximation capabilities of multilayer feedforward networks," *Neural networks*, vol. 4, no. 2, pp. 251-257, 1991.
133. S. Ameer, M.A. Shah, A. Khan, H. Song, C. Maple et al., "Comparative analysis of machine learning techniques for predicting air quality in smart cities," *IEEE Access*, vol. 7, no. pp. 128325-128338, 2019.
134. X. Liu and H. Guo, "Air quality indicators and aqi prediction coupling long-short term memory (lstm) and sparrow search algorithm (ssa): A case study of shanghai," *Atmospheric Pollution Research*, vol. 13, no. 10, pp. 101551, 2022.
135. C. Suresh, B. Kiranmayee and B. Sneha. "Analysis and prediction of air pollutant using machine learning". in *Proceedings of Second International Conference on Advances in Computer Engineering and Communication Systems: ICACECS 2021*. 2022. Springer.
136. H. Nguyen, M. Mehrabi, B. Kalantar, H. Moayedi and M.a.M. Abdullahi, "Potential of hybrid evolutionary approaches for assessment of geo-hazard landslide susceptibility mapping," *Geomatics, Natural Hazards and Risk*, vol. 10, no. 1, pp. 1667-1693, 2019.
137. H. Moayedi, M. Mehrabi, M. Mosallanezhad, A.S.A. Rashid and B. Pradhan, "Modification of landslide susceptibility mapping using optimized pso-ann technique," *Engineering with Computers*, vol. 35, no. 3, pp. 967-984, 2019.
138. H. Abedinpourshotorban, S.M. Shamsuddin, Z. Beheshti and D.N. Jawawi, "Electromagnetic field optimization: A physics-inspired metaheuristic optimization algorithm," *Swarm and Evolutionary Computation*, vol. 26, no. pp. 8-22, 2016.

139. N. Van Thieu, S.D. Barma, T. Van Lam, O. Kisi and A. Mahesha, "Groundwater level modeling using augmented artificial ecosystem optimization," *Journal of Hydrology*, vol. 617, no. pp. 129034, 2023.
140. H. Moayedi and A. Mosavi, "An innovative metaheuristic strategy for solar energy management through a neural networks framework," *Energies*, vol. 14, no. 4, pp. 1196, 2021.
141. I. Aranguren, A. Valdivia, M. Pérez-Cisneros, D. Oliva and V. Osuna-Enciso, "Digital image thresholding by using a lateral inhibition 2d histogram and a mutated electromagnetic field optimization," *Multimedia Tools and Applications*, vol. 81, no. 7, pp. 10023-10049, 2022.
142. X. Feng, Q. Li, Y. Zhu, J. Hou, L. Jin et al., "Artificial neural networks forecasting of pm2. 5 pollution using air mass trajectory based geographic model and wavelet transformation," *Atmospheric Environment*, vol. 107, no. pp. 118-128, 2015.
143. F. Biancofiore, M. Busilacchio, M. Verdecchia, B. Tomassetti, E. Aruffo et al., "Recursive neural network model for analysis and forecast of pm10 and pm2. 5," *Atmospheric Pollution Research*, vol. 8, no. 4, pp. 652-659, 2017.
144. A. Azid, H. Juahir, M.E. Toriman, M.K.A. Kamarudin, A.S.M. Saudi et al., "Prediction of the level of air pollution using principal component analysis and artificial neural network techniques: A case study in malaysia," *Water, Air, & Soil Pollution*, vol. 225, no. pp. 1-14, 2014.
145. H. Abdi and L.J. Williams, "Principal component analysis," *Wiley interdisciplinary reviews: computational statistics*, vol. 2, no. 4, pp. 433-459, 2010.
146. J.-O. Kim, O. Ahtola, P.E. Spector, J.-O. Kim and C.W. Mueller, "Introduction to factor analysis: What it is and how to do it," in ed., Place, Sage, 1978.
147. H.F. Kaiser, "The varimax criterion for analytic rotation in factor analysis," *Psychometrika*, vol. 23, no. 3, pp. 187-200, 1958.
148. C.-W. Liu, K.-H. Lin and Y.-M. Kuo, "Application of factor analysis in the assessment of groundwater quality in a blackfoot disease area in taiwan," *Science of The Total Environment*, vol. 313, no. 1-3, pp. 77-89, 2003.
149. E. Sarwat and G.I. El-Shanshoury, "Estimation of air quality index by merging neural network with principal component analysis," *International Journal of Computer Application*, vol. 1, no. 8, pp. 2250-1797, 2018.
150. B. Wang, H. Moayedi, H. Nguyen, L.K. Foong and A.S.A. Rashid, "Feasibility of a novel predictive technique based on artificial neural network optimized with particle swarm optimization estimating pullout bearing capacity of helical piles," *Engineering with Computers*, vol. 36, no. pp. 1315-1324, 2020.
151. J.J. Moré, "*The levenberg-marquardt algorithm: Implementation and theory*", in *Numerical analysis*. 1978, Springer. p. 105-116.
152. O. Asadi Nalivan, S.A. Mousavi Tayebi, M. Mehrabi, H. Ghasemieh and M. Scaioni, "A hybrid intelligent model for spatial analysis of groundwater potential around urmia lake, iran," *Stochastic Environmental Research and Risk Assessment*, vol. no. pp. 1-18, 2022.
153. H. Moayedi, S. Ghareh and L.K. Foong, "Quick integrative optimizers for minimizing the error of neural computing in pan evaporation modeling," *Engineering with Computers*, vol. no. pp. 1-17, 2021.
154. L.K. Foong, Y. Zhao, C. Bai and C. Xu, "Efficient metaheuristic-retrofitted techniques for concrete slump simulation," *Smart Structures and Systems, An International Journal*, vol. 27, no. 5, pp. 745-759, 2021.
155. C. Lin and J. Wang, "Metaheuristic-designed systems for simultaneous simulation of thermal loads of building," *Smart Structures and Systems*, vol. 29, no. 5, pp. 677-691, 2022.
156. J.-S. Jang, "Anfis: Adaptive-network-based fuzzy inference system," *IEEE transactions on systems, man, and cybernetics*, vol. 23, no. 3, pp. 665-685, 1993.

157. M. Mehrabi, B. Pradhan, H. Moayedi and A. Alamri, "Optimizing an adaptive neuro-fuzzy inference system for spatial prediction of landslide susceptibility using four state-of-the-art metaheuristic techniques," *Sensors*, vol. 20, no. 6, pp. 1723, 2020.
158. H. Moayedi, M. Mehrabi, B. Kalantar, M. Abdullahi Mu'azu, A.S. A. Rashid et al., "Novel hybrids of adaptive neuro-fuzzy inference system (anfis) with several metaheuristic algorithms for spatial susceptibility assessment of seismic-induced landslide," *Geomatics, Natural Hazards and Risk*, vol. 10, no. 1, pp. 1879-1911, 2019.
159. A. Yildirim, M. Bilgili and A. Ozbek, "One-hour-ahead solar radiation forecasting by mlp, lstm, and anfis approaches," *Meteorology and Atmospheric Physics*, vol. 135, no. 1, pp. 10, 2023.
160. M. Mehrabi, "Landslide susceptibility zonation using statistical and machine learning approaches in northern lecco, italy," *Natural Hazards*, vol. no. pp. 1-37, 2021.