HUNGARIAN UNIVERSITY OF AGRICULTURE AND LIFE SCIENCES

THE VARIOUS TRANSMISSION AND ESTIMATION OF PM POLLUTANTS

Thesis of the doctoral (PhD) dissertation

Qor-el-aine Achraf

Gödöllő - Hungary
2023

Doctoral school denomination: Doctoral School of Mechanical Engineering

Science: Mechanical Engineering – Environmental Engineering

Head of school: Prof. Dr. Gábor Kalácska, DSc
Institute of Technology
Hungarian University of Agriculture and Life Science, Gödöllő - Hungary

Supervisor: Dr. Gábor Géczi, PhD
Institute of Environmental Sciences
Hungarian University of Agriculture and Life Science, Gödöllő - Hungary

Co-Supervisor: Dr. András Béres, PhD
University Laboratory Center, Head of Szent István Campus
Hungarian University of Agriculture and Life Science, Gödöllő – Hungary

…… ……

Affirmation of head of school Affirmation of supervisor

CONTENTS

LIST OF SYMBOLS

DS	concentration Dust surface particles	[µg.m^{-3}]
DS2.5	concentration Dust surface particles with an aerodynamic diameter of less than 2.5 µm	[µg.m^{-3}]
BC	Black carbon concentration	[µg.m^{-3}]
OC	Organic Carbon concentration	[µg.m^{-3}]
SS	Sea salt particle concentration	[µg.m^{-3}]
SS2.5	concentration of Sea salt particle with an aerodynamic diameter of less than 2.5 µm	[µg.m^{-3}]
NO_2	Nitrogen dioxide concentration	[µg.m^{-3}]
O_3	Ozone concentration	[µg.m^{-3}]
P	pressure at ground surface	[kPa]
PBLH	Planetary boundary layer height	[m]
PM10	Concentration of particles with an aerodynamic diameter up to 10 µm	[µg.m^{-3}]
PM2.5	Concentration of particles with an aerodynamic diameter of less than 2.5 µm	[µg.m^{-3}]
RH	relative humidity	[-]
SO_2	Sulphur dioxide concentration	[µg.m^{-3}]
SO_4	Sulphate concentration	[µg.m^{-3}]
T	temperature at 2 m of the ground surface	[°C]
Ws_{10}	Wind speed at 10 m of the ground surface	[m.s^{-1}]
Ws_{50}	Wind speed at 50 m of the ground surface	[m.s^{-1}]

Abbreviations

AOD	Aerosol Optical Depth
AERONET	AErosol RObotic NETwork
CALIPSO	The Cloud-Aerosol Lidar and Infrared Pathfinder Satellite Observation
CAMS	The Copernicus Atmosphere Monitoring Service
HYSPLIT	Hybrid Single-Particle Lagrangian Integrated Trajectory
MAE	Mean Absolute Error
MERRA	Modern-Era Retrospective analysis for Research and Applications
MLR	Multiple Linear Regression
MODIS	The Moderate Resolution Imaging Spectroradiometer
VIIRS	The Visible Infrared Imaging Radiometer Suite
SDE	Saharan Dust Event
OLS	Ordinary Least Square regression
PM	Particulate Matter
RMSE	Root of Mean Square Error
WHO	World Health Organization

1 INTRODUCTION

The context and goals of this PhD thesis are described in this chapter.

1.1 Relevance and significance of the topic

Air pollution is the primary cause of the decline in air and environmental quality in many places across the world nowadays, with negative consequences for people's health. According to the most recent World Health Organization (WHO) report, more than 91% of people in urban areas are exposed to air quality levels that exceed the emission limits for air pollution (World Health Organization, 2021). Carbon monoxide (CO), particulate matter (PM), nitrogen oxides (NOx), volatile organic compounds (VOCs), ozone (O_3), and sulphur dioxide (SO_2) are the primary atmospheric pollutants. The rapid industrialization and urbanization of developing countries have increased the number of pollutants emitted (Fu and Chen, 2017). Because of the strong relationship between air pollution exposure and increased harmful short- and long-term effects on human health, the scientific community, and public opinion are both concerned about the deterioration of air quality in urban environments (Masiol *et al.*, 2014). Aside from the health dangers posed by gas and particle inhalation, urban air pollution causes other issues such as faster corrosion and deterioration of materials, damage to historical monuments and structures, and damage to plants in and around the city (Vlachokostas *et al.*, 2011).

Particulate matter (PM) is a broad word that refers to a mixture of solid particles and liquid droplets (aerosols) whose size and composition change depending on time and place. PM is composed of numerous constituents, including elemental or black carbon (BC) and organic carbon (OC) molecules, sulfate (SO_4^{-2}), nitrate (NO_{3-}), trace metals, soil particles, and sea salt. PM particles are defined based on their size variations. PM particles with a diameter that is less than or equal to 10 µm are called coarse PMs (PM10), and PM with a diameter of less than or equal to 2.5 µm are fine PMs (PM2.5). PM can be directly emitted from anthropogenic or natural sources (i.e., primary PM), or formed in the atmosphere from chemical reactions of numerous gaseous (i.e., secondary PM) (Harrison, Hester and Querol, 2016).

The research of PM pollution is crucial, in order to comprehend the causes and effects of this kind of air pollution and to create practical solutions for lowering exposure and enhancing public health. Researchers employ a range of techniques to evaluate PM pollution, including computer modelling, satellite data, and air quality monitoring stations. These techniques can offer details on the concentration and distribution of PM in various locations as well as the pollution's origins. Chemical analysis, for instance, may be used to determine the chemical composition of PM samples and link them to particular sources such industrial activities, wildfires, or vehicle emissions.

This research, covers different aspects of the PM pollution, from the evaluation of low-cost PM sensors, the use of low-cost PM sensors in small scale experiments, the effect of dust storms on the PM concentrations in Hungary and how often they occur, and the estimation of the PM10 and PM2.5 concentrations based on Satellite, meteorological and in-situ measurements data.

1.2 Objectives

The primary goals of the present work are the following:

1) Study the effects of a simple environment on PM concentration using PM low-cost sensors.
2) Study the effects of the Saharan Dust storm on PM levels in Hungary and analyse the seasonality and frequency of recent Saharan dust events.
3) Estimating PM concentrations using satellite, meteorological, and in-situ measurement data and machine learning methods over Hungary.

2 LITERATURE REVIEW

2.1 Measurement of PM

Many approaches have been employed to estimate, forecast, or measure PM concentrations. Ambient monitoring at stationary sites offers consistent concentration data that can frequently span decades. However, monitoring stations may be rare in regional regions due to cost and facility constraints. While Satellite remote sensing has been used successfully to estimate PM concentrations in many regions around the world to overcome the shortcoming of the geographical cover of fixed monitoring air quality stations.

2.1.1 In Situ Measurements

The determination of PM concentrations in ambient air is complicated and highly dependent on the approaches utilized. Mass concentration or particle number are two factors that may be determined as a function of particle size. In terms of measuring PM mass, filter-based devices can be employed to collect PM for later gravimetric or automatic determination. The first approach requires weighing the filters "offline," but the second method gives an indirect assessment of PM mass online. Automated analysers can be based on β-ray attenuation, a glass tapered tube's frequency of oscillation, or particle light scattering (nephelometer). The filter-based devices generate a sample that may be utilized for PM chemical analysis (Minguillón, Viana and Querol, 2013).

The filter-based gravimetric technique is the international technical standard for detecting particle mass concentrations. This approach is based on weighing the PM accumulated on a filter within a sampler. An analytical balance is used to weigh the filter before and after sampling. Weighing occurs after the filter has been kept under regulated humidity and temperature conditions for a certain amount of time. The difference between the two weighing represents the mass of PM collected, which is then divided by the volume of air passed through the filter to get the mass concentration of PM. Furthermore, field mass measurements can be performed using automated analysers. One of the most extensively used methods for assessing ambient PM concentrations is the analyser based on the β-ray attenuation concept. According to the Beer-Lambert relationship, β-radiation flowing through the filter sample is attenuated by the particle collected on the filter, and PM mass is proportional to the degree of attenuation. As a result, the difference in attenuation before and after the sample period offers an indirect estimate of PM mass concentration. These devices are intended to allow for a filter replacement every 24 hours, eliminating the need for daily service (Baron, Kulkarni and Willeke, 2011).

In air quality monitoring stations, beta attenuation is commonly utilized (Liu *et al.*, 2022). Cheng *et al.* (2016) studied the PM2.5 pollution status of 45 megacities throughout the world, and 28 of the 45 cities employ β-ray attenuation to assess particulate matter concentrations in the atmosphere. The β-ray attenuation technique offers a high measurement accuracy and a faster reaction time than the filter weighing method, often up to 1 hour, and it is widely employed in air quality monitoring stations in many countries (Liu *et al.*, 2022).

The European Union introduced ambient air quality directives to enhance air quality by evaluating the level of air pollutants in ambient air across Europe, as well as to safeguard the environment and human health. The PM10 limit values according to EU standards are as follows: a yearly average of 40 $\mu g/m^3$, and a daily limit value of 50 $\mu g/m^3$ may not be surpassed over than 35 days of the year. While for PM2.5 there is no daily limit concentration, and the yearly limit value is 25 $\mu g/m^3$ (Brunekreef and Maynard, 2008).

In Europe, PM is still the most threatening air pollutant to human health, due to its serious effects on health (Guerreiro, Foltescu and de Leeuw, 2014). According to the latest report of Air quality

in Europe (European Environment Agency, 2022), Italy and numerous eastern European nations have exceeded the EU daily maximum threshold for PM10, and registering high PM2.5 concentrations. Solid fuels like coal are commonly utilized for heating residences in most Central and Eastern European nations, as well as in certain industrial buildings and power plants. The Po Valley, located in northern Italy, is a heavily inhabited and industrialized area with unique climatic and geological circumstances that favour the buildup of air pollutants in the atmosphere.

In Hungary, many cities suffer from high PM concentrations, and the most important factors influencing the PM concentrations in Hungary are local anthropogenic emissions like traffic and biomass burning, in addition to long range transport of particles from outside the country and the combination of the previous two factors with meteorological conditions (Ferenczi and Bozó, 2017; Perrone *et al.*, 2018). The reports of air quality in Europe throughout the year show high concentrations of PM10 and PM2.5 especially in Northern and Central Hungary. The latest report of Air quality in Europe (European Environment Agency, 2022), show high concentrations of PM10 in Northern Hungary (Miskolc, Nyíregyháza, Kazincbarcika and Sajószentpéter), in Central Hungary (Budapest and Dunaújváros), and in Southern Transdanubia region (Pécs), while for PM2.5 the report show that the concentrations are high in all Hungarian air quality stations that report measurements to the EU (Figure 2.1 and 2.2). Smog was observed in Hungary around the end of January 2017. Across the country, PM10 readings were over the limit. Extremely high concentrations were recorded in the eastern part of Hungary on January 31, 2017. During this time period, the peak one-hour average concentration was above 500 µg/m^3 in the Sajó Valley, which is situated in northern Hungary (Ferenczi, Homolya and Bozó, 2020).

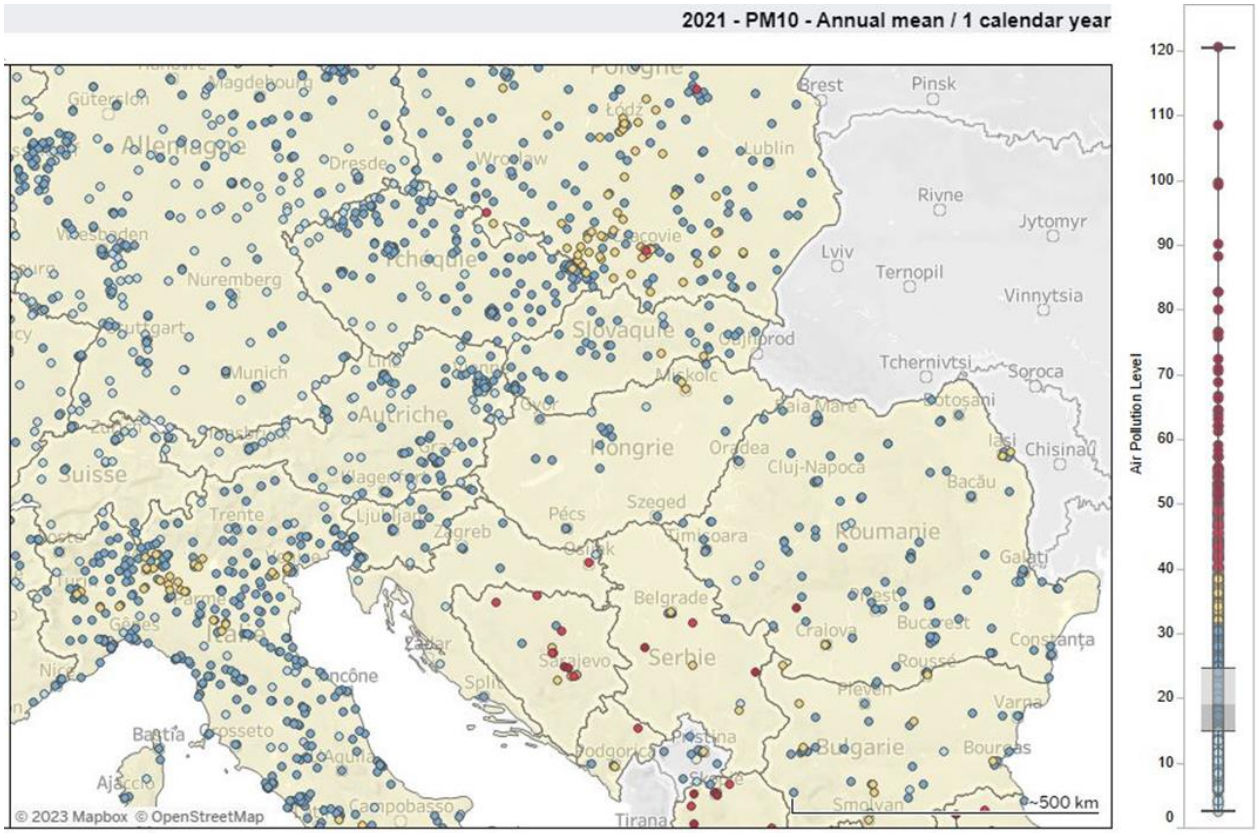

Figure 2.1: Annual mean PM10 map concentrations in Hungary in 2021 (European Environment Agency., 2022)

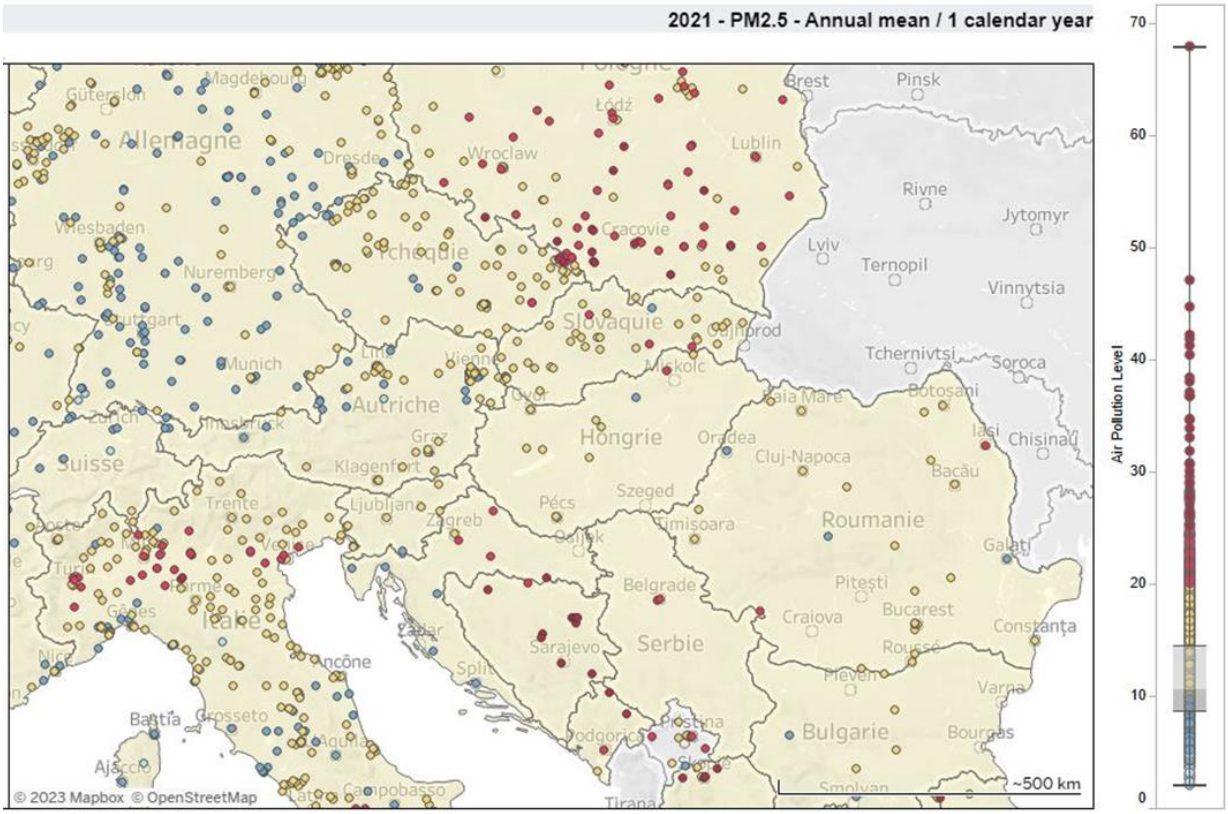

Figure 2.2: Annual mean PM2.5 map concentrations in Hungary in 2021 (European Environment Agency., 2022)

2.1.2 Low-cost PM sensors

Low-cost sensors do have ability to significantly change how, where, and when air pollution is monitored. Until around 2010, routine measurements of particulate matter mass (PM) concentrations can only be undertaken by entities that could afford instruments in the tens of thousands of dollars, with entire station and staff expenses substantially an order of magnitude greater. Over the last decade, there have been significant changes in how air quality monitoring is conducted, who may undertake it, and the temporal and geographical dimensions on which we look regarding air quality in general. The advent of low-cost air quality sensors, notably low-cost PM monitors, has fuelled these shifts (Giordano *et al.*, 2021). The major advantage of low-cost PM sensors is obvious from the name; these devices are affordable. Most PM low-cost sensor cost between $10 and $100 for a single sensor, while the accompanying electronics and packaging may raise the price to between $200 and $2000 - still a low fraction of the price of typical reference PM monitors (Giordano *et al.*, 2021). Light-scattering is the operating principle of approximately all low-cost PM sensors available on the market. Light scattering has a substantially lower form factor than most other particle counting/mass concentration measuring techniques. A low-cost PM sensor has only three essential components: a light emitting diode (usually infrared or red laser), a phototransistor, and a lens to concentrate the diode light. To pull air into the measuring cavity, a fan, pump, or convective heater might be utilized (Wang *et al.*, 2015). The measuring concepts of low-cost PM sensors are straightforward: when particles pass through into the measuring cavity, the intensity of light of the infrared/red light entering the phototransistor is altered by the presence of particles. The strength of the fluctuation, i.e., the nephelometric response, is proportional to particle concentration (mass and number). The optics (diode, transistor, and lens) and shape of the measuring cell have a significant influence on the total nephelometric response of these devices. As a result, even if the measuring principles are the similar across low-cost PM sensors, different

types and models of low-cost sensors frequently behave fairly differently (Williams *et al.*, 2015). One of the hallmarks of low-cost PM sensors is that they frequently assess scattering according to a particle population rather than individual particle scattering. This group measurement mitigates some of the disadvantages of monitoring individual particles via light scattering. The low-pulse occupancy approach is used in many low-cost sensors: the phototransistor outputs either a high voltage (when the light channel is clean) or a low voltage (when the light path includes particles) in a modulating pulse (Giordano *et al.*, 2021).

Some of the low-cost PM sensors have already been tested in the outdoors and in laboratories. These experiments demonstrated that these PM sensors have a lot of potential (Kelly *et al.*, 2017). Wang *et al.* (2015) conducted controlled laboratory investigations on three low-cost sensors and discovered that their PM measures corresponded linearly ($R^2 = 0.89$) with research-grade devices. Austin *et al.* (2015) conducted laboratory tests and discovered a linear correlation ($R^2 = 0.66$-0.99) between the low-cost PM sensors used and a research-grade device. Aerosol particle sizer over a concentration range of 1-50 µg/m3, though the slope of the linear relationship varied by more than a factor of 10 depending on particle diameter. In a 4-day investigation, Gao, Cao and Seto, (2015) examined the responsiveness of low-cost PM sensor in a polluted region of China where concentration of PM2.5 in 24h period ranged between 330-413 µg/m^3, and observed correlations to compile research-grade instruments ($R^2 = 0.86$-0.89) and gravimetric measurements ($R^2 = 0.53$). Holstius *et al.* (2014) examined the performance of low-cost PM sensors to 1h and 24h PM2.5 data out of a class III US EPA Federal Equivalent Method, measured by PM2.5 β-attenuation monitor in continuous operation at an Oakland, California regulatory monitoring site. They discovered strong correlation that explained 60% of the variation in 1 hour reference PM2.5 data and 72% of the variance in 24-hour data. Liu *et al.* (2020) also investigated the influence of temperature and relative humidity on sensor performance, the responses of PM2.5 sensors to different kinds of aerosols, and long-term stability at four locations in Australia and China. For PM2.5 readings, the PM2.5 sensors had good inter-correlations ($r > 0.91$). The performance of the PM2.5 sensors varied depending on location, with moderate to strong correlations with reference equipment (R^2 was between 0.44 and 0.91). With 75% relative humidity and high temperatures, sensors operated admirably. The PM2.5 sensor was more sensitive to mixed urban background pollutants, aged traffic emissions, and industrial emissions than to sea aerosols and fresh vehicle emissions. The PM2.5 sensors have decent long-term stability. Malings *et al.* (2020) tested two low-cost PM2.5 sensors over a lengthy period of time. In the short term, they discovered that both sensors produced relatively precise concentration measurements with a mean absolute error of approximately 4 µg/m^3 in near-real time. The yearly average error was decreased to less than 1 µg/m^3 for the long-term scenario, which was evaluated through year-long collocations at one urban background and one near-source site. The conclusion was that low-cost PM2.5 sensors may be used to complement scarce networks of regulatory-grade devices, undertake high-density nearby monitoring, and help researchers better comprehend spatial variation and temporal air quality changes in urban areas. In addition, Báthory *et al.* (2019) evaluated a low-cost dust sensor (Plantower PMS7003) implemented in stationary and mobile PM measurements devices in Miskolc Hungary. The results were that the mobile device may tour a region and find the areas with the highest concentrations. Also, Báthory *et al.* (2022) tested the performance of a low-cost sensor (Plantower PMS7003) calibrated in a controlled climatic chamber against a conventional reference aerosol monitor. The study's findings showed remarkable unit-to-unit consistency; nevertheless, each sensor had to be calibrated separately because their features differed significantly. According to the findings of a 15-month field test, quantitative and indicative low-cost sensor performance seemed promising: overall indicative accuracy was roughly 73-75% with comparable precision and recall, and it is recommended that the low-cost sensor be cleaned after 6-8 months of operation.

While low-cost PM sensors are becoming more popular due to their price and accessibility, they have significant limitations that must be considered when interpreting their data and making informed judgments about air quality (Raysoni *et al.*, 2023). One of the most noticeable drawbacks is their lower accuracy and precision than reference-grade devices, which frequently provide approximate results that can be altered by factors such as sensor calibration, sensor degradation, and cross-sensitivity to diverse environmental conditions (Raysoni *et al.*, 2023). Regular sensor calibration is necessary, but it can be complicated and time-consuming, and the quality of calibration standards might vary, reducing the trustworthiness of the sensor's data. These sensors can also drift over time, resulting in less accurate results and the need for regular recalibration or maintenance (Karagulian *et al.*, 2015). Furthermore, many low-cost sensors have a restricted measuring range, making them inappropriate for extraordinarily high or low particulate matter concentrations. They are susceptible to environmental conditions, including temperature, humidity, and altitude, which can affect their readings and require compensation or correction to maintain accuracy (Alfano *et al.*, 2020). Furthermore, unlike reference devices that can measure numerous particle sizes and provide more precise information about air quality, many of these sensors report only a single PM value, often PM2.5 or PM10, without discriminating between different particle sizes (Giordano *et al.*, 2021). They may also lack the ability to distinguish between different types of particles, reporting high PM readings when other particles in the air, such as dust or pollen, are present (Oluwadairo *et al.*, 2022). Because low-cost sensors are not subject to the same stringent quality control and performance standards as regulatory-grade devices employed by government organisations, sensor quality and dependability can vary significantly between manufacturers and models (Liu *et al.*, 2020). Proper data validation and comparison to reference instruments are required to ensure accuracy, but without these steps, questions about the measurement's dependability may develop (Alfano *et al.*, 2020). Furthermore, low-cost sensors may have a shorter operating lifespan than reference instruments, necessitating frequent replacement or maintenance increasing overall costs (Karagulian *et al.*, 2015). Integrating their data into current air quality monitoring networks might be difficult because of their non-standardized nature, making it impossible to compile comprehensive, standardised information for a region (Giordano *et al.*, 2021). Finally, analysing data from low-cost sensors can be complex for the general public because it frequently necessitates a grasp of the technology's limits and potential biases (Sayahi, Butterfield and Kelly, 2019). Despite these limitations, low-cost PM sensors can still be valuable tools for raising awareness about air quality, conducting community-based monitoring, and identifying air quality trends and pollution hotspots when used with caution, alongside reference-grade instruments, and with the understanding that their results should be regarded as indicative rather than definitive measurements of air quality.

2.1.3 Ambient PM pollution

Ambient PM, or outdoor PM pollution, can come from a variety of sources, including natural and anthropogenic sources. Some common sources of ambient PM particles include:

- Transportation: Automobiles are a significant generator of ambient PM, particularly in cities. PM particles such as soot, carbon monoxide, and nitrogen oxides can be found in exhaust emissions from vehicles, trucks, and buses.
- Residential Heating and Cooking: covers the use of wood, coal, and gas for cooking or heating. Central European region is typical example of household PM pollution due to the use of wood and coal for domestic heating.
- Agricultural activities: Farming and livestock production can produce PM particles such as dust from plowing and harvesting, as well as ammonia from animal waste.
- Natural sources of PM: include sea salt, wind-blown dust (Saharan dust), wildfires, and volcanic eruptions.

- Industries: Significant sources of PM emissions include manufacturing, mining, and construction as well as the burning of fossil fuels for energy generation.
- Secondary PM particles: secondary PM particles are produced as a result of chemical reactions in the atmosphere. Secondary aerosols are created when major PM precursors (such as sulfur dioxide, nitrogen oxides, and volatile organic compounds) mix with other atmospheric components (such as water vapor, oxygen, and sunlight) to form secondary aerosols.

Ammonium sulfate, ammonium nitrate, organic aerosols, and secondary organic aerosols are examples of secondary PM particles. The reactions of sulfur dioxide and nitrogen oxides with ammonia produce ammonium sulfate and ammonium nitrate, whereas organic aerosols and secondary organic aerosols are produced by the oxidation of organic molecules (Seinfeld and Pandis, 2016).

In an extensive study Thunis *et al.* (2018) evaluated 150 urban regions in Europe to determine source allocation of PM2.5 using SHERPA modelling tool which is designed to investigate prospective air quality improvements as a result of national/regional/local emission reduction measures. The results were the followings:

- The residential sector contributes 13% on average. All of the major contributions were found in Poland, with the maximal values in Warsaw (48%), Krakow (40%), Katowice (40%), Lodz (33%), and Poznan (33%). Generally, the impact of household heating is greater in eastern European countries (particularly Poland) and in some Italian cities.
- Traffic contributes 14% on average. 39%, 30%, 29%, 27%, and 27%, were the highest contributions values found in Madrid, Luxembourg, Paris, Verona, and Bologna respectively. As predicted, transportation emissions play a significant role in several of the world's major cities (Paris, Madrid, London). They do, however, play an important role in heavily populated countries such as Belgium and the Netherlands.
- Agriculture contributes an average of 23%. Dresden (40%), Braunschweig-Salzgitter-Wolfsburg (39%), Usti nad Labem (38%), Plzen (37%), and Leipzig (36%), are the cities with the highest contributions. Agricultural activity and associated emissions are not often conducted within urban limits, while certain agricultural regions may exist within larger metropolitan areas. Yet, the sector contributes significantly to PM2.5 concentrations (secondary pollution) in several EU cities, particularly in central Europe.
- Industry contributes 20% on average. Mannheim-Ludwigshafen (47%), Bilbao (46%), Linz (44%), Marseille (41%), and Brescia (37%) have the highest donations. Industry is important in several Eastern nations (Bulgaria, Romania, and Greece), as well as the west side of Germany. It is less important in southern Europe, despite being a major contributor in places such as Marseille and Turin.
- Natural sources account for an average of 19%. PM2.5 concentration maxima in Mediterranean cities are linked to Saharan dust event occurrences. Valletta (46%), Limassol (43%), Palma de Mallorca (40%), Nicosia (39%), and Alicante (36%), have the highest contributions.

Moreover, indoor PM pollution can also contribute to outdoor PM pollution, cooking and heating with biomass fuels also contribute significantly to outdoor air pollution, with an estimated 10% of ambient PM2.5 (Diette *et al.*, 2012). Household sources can account for substantial amount of PM emission to ambient atmosphere. Small-scale wood burning, which occurs mostly throughout Europe, Africa, and Asia, is responsible for a significant amount of PM in the atmosphere (Butt *et al.*, 2016). Almost 20% of the total mass of PM2.5 is produced worldwide as a result of residential fuel burning. For Central and Eastern Europe, North-western Europe, Western Europe, and South-

western Europe, the corresponding percentage contributions are 32%, 22%, 15%, and 12%, respectively (Karagulian *et al.*, 2015).

Ambient PM pollution, has been related to a variety of acute and chronic health impacts. Because of its capacity to penetrate deep into the lungs and even reach the bloodstream, PM2.5 has been identified as a key contributor to these effects.

Because of their high penetrability, PM particles can move through the respiratory system from the nose passages to the alveoli deep within the lungs. Particles with diameters of 5 to 10 µm are most often deposited in the tracheobronchial tree, whereas those with diameters of 1 to 5 µm typically deposited in the respiratory bronchioles and alveoli, where gas exchange occurs (Löndahl *et al.*, 2006) (Figure 2.3). These particles can interfere with gas exchange in the lungs and possibly infiltrate the lung. These particles will eventually enter the bloodstream and create serious health concerns. Particles smaller than 1 µm act similarly to gas molecules and, as a result, will enter the alveoli and can translocate deeper into the cell tissue and/or circulatory system (Valavanidis, Fiotakis and Vlachogianni, 2008).

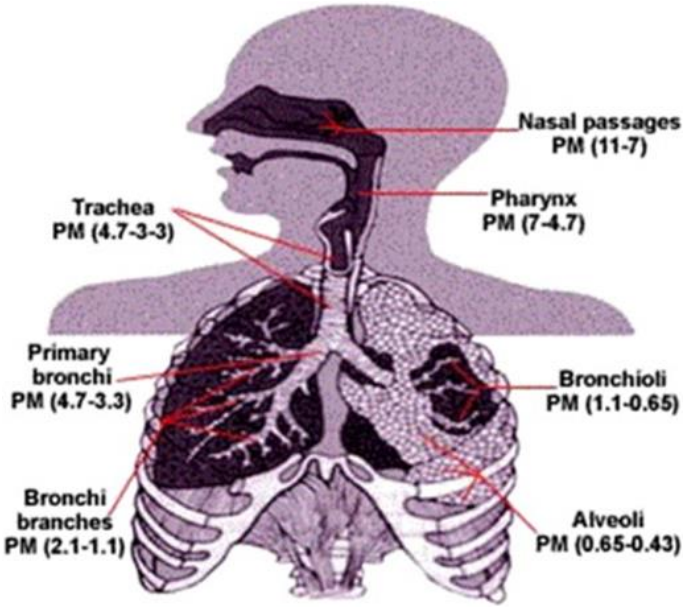

Figure 2.3: Possibility of particle deposition in a variety of sizes (Kim, Kabir and Kabir, 2015)

Furthermore, experts believe that exposure to high PM concentrations may cause a variety of symptoms, including low birth weight in newborns, pre-term births, and perhaps infant fatalities (Fong *et al.*, 2019). Minor side effects of breathing PM2.5 include shortness of breath, chest discomfort and soreness, and coughing and wheezing (Guaita *et al.*, 2011). When kids are subjected to excessive levels of PM2.5 (for instance 65 µg/m^3 for a day), they are at a considerably increased danger for respiratory symptoms, the use of asthma medications, and impaired lung function (Guaita *et al.*, 2011). According to Sofer *et al.* (2013), individuals exposed to high PM10 levels (higher than 150 µg/m^3) had a 3 to 6% deterioration in lung function as determined by maximal expiratory flow. Tecer *et al.* (2008) observed an 18% increase in asthma patient admissions with a 10 µg/m^3 increase in PM10 during the same day's admissions. Panel research in Los Angeles reported that a 10 µg/m^3 rise in PM2.5 was related with an increase in carotid intima-media thickness, an ultrasonic marker of atheroma, in 798 individuals in two clinical trials (Künzli *et al.*, 2005)

Exposure to PM2.5 for a few hours to a few weeks can cause cardiovascular disease-related mortality and nonfatal events; repeated exposure raises the risk for cardiovascular mortality to

even larger extent than exposure for a few days and reduces life expectancy in more highly exposed segments of the population by several months to a few years; and declines in PM levels are associated with reduces in cardiovascular mortality within as little as a few weeks. The overall evidence supports a causal connection among PM2.5 exposure and cardiovascular disease and death (Brook *et al.*, 2010). PM is thought to be responsible for around 3% of cardiopulmonary and 5% of lung cancer deaths worldwide (Fang *et al.*, 2013). Exposure to PM2.5 is expected to lower the population's life expectancy by around 8.6 months on average (Krewski, 2009). Based on data sets gathered from 545 counties in the United States from 2000 to 2007, Correia *et al.* (2013) proposed a probable relationship between fine particulate matter decrease and enhanced life expectancy. According to study findings, a reduction of 10 $\mu g/m^3$ of PM2.5 ought to have resulted in an improvement in life expectancy of 0.35 years in average. In a recent review study by Chen *et al.* (2022), they showed other aspects of the health effects of PM. They emphasise that Exposure to PM leads to disorders in multiple organ systems, including the lungs, heart, liver, and kidneys, as well as oxidative stress and inflammation, which are common mechanisms involved in the adverse health effects of PM exposure. Shi *et al.* (2023) concluded that exposure to PM concentrations, even at concentrations below the air quality threshold in the EU, was associated with the incidence and Mortality of pulmonary hypertension and that the probability of the transition from pulmonary hypertension to death increased due to exposure of various ambient air pollutants including PM10 and PM2.5. Also, being exposed to PM2.5 can have effects on the dietary intake. Sundram *et al.* (2022) showed that long hours of PM2.5 exposure affect personal dietary intake, potentially increasing the risk of metabolic syndromes and other undesired health conditions. The study conducted a 13-week follow-up involving two cohorts of outdoor and indoor workers. The results showed that outdoor workers exposed to high levels of PM2.5 had significant changes in appetite and increased calorie intake. Another aspect of the effects of PM pollutants on humans is the increased risk of obesity. In a meta-analysis of the global association between atmospheric PM and obesity, Lin *et al.* (2022) highlight that exposure to PM, particularly PM2.5, increases the risk of obesity, especially in Asian populations. Maternal exposure to PM2.5 is found to aggravate obesity in children. The study also indicates that minors are more susceptible to obesity with PM2.5 and PM10 exposure. Depending on the PM2.5 particle's chemical composition, it can trigger an inflammatory response in the body where this chronic inflammation has been linked to the development of obesity and insulin resistance.

2.2 Saharan dust and PM

Natural resource dust storms and wildfires may be significant sources of suspended dust. Globally, arid and semiarid regions are responsible for the majority of dust emissions. The dust belt (Figure 2.4) consists of primary dust source regions that stretch from the west coast of North Africa, the Middle East, central and south Asia to China (Ashrafi *et al.*, 2014).

Sahara is the major and most active dust source in the world. The approximate emission ranges from 400 to 2200 $Tg.yr^{-1}$, which represent 36% to 79% of world emissions (Huneeus *et al.*, 2011). The Saharan dust contains natural mineral particles, that by interaction with anthropogenic pollutants can exacerbate secondary aerosol formation (Ashrafi *et al.*, 2014). Mineral dust also contains elements like iron and phosphorous, which are important nutrients in both terrestrial and marine ecosystems. Mineral dust deposited in the ocean, for example, might impact primary productivity, which in turn can affect the atmosphere-ocean carbon cycle and, eventually, climate (Ravi *et al.*, 2011; Jickells and Moore, 2015). There has long been an interest in the African dust. Darwin was on the Beagle when a violent dust storm hit the Cape Verde Islands in 1833, that he describes in is extensive measurements published in 1846 (Darwin, 1846), finishing with a comment that he has no doubt that the dust which fell in the Atlantic came from Africa.

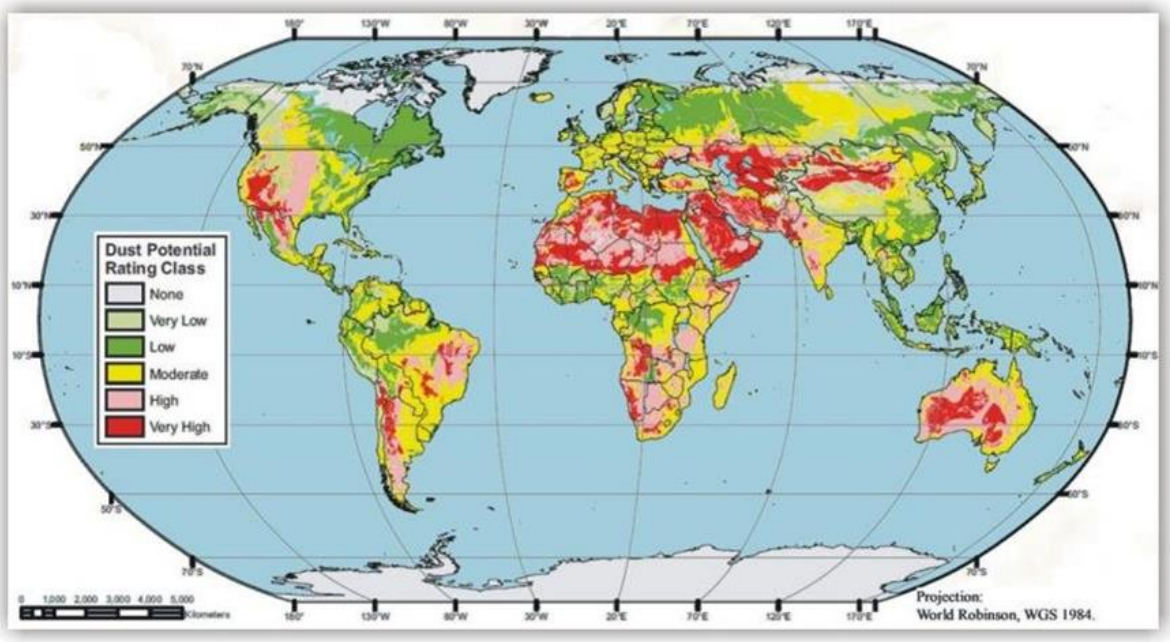

Figure 2.4: Global Dust Potential source (Shepherd *et al.*, 2016)

Dust storms are atmospheric phenomena caused by the erosion and movement of mineral deposits from the earth. The process of erosion and transported dust from ground surface is divided into three stages: surface material entrainment or emission, atmospheric transport, and deposition. Particle entrainment occurs when the wind shear force applied on the surface (wind erosivity) surpasses the surface material's capacity to resist separation or transport (sediment or soil erodibility). Particle movement can be caused by wind transport via the processes of saltation, creep, and suspension (Figure 2.5). Particles with diameters greater than 500 microns will slide across the earth surface. Saltation occurs when wind moves particles ranging in size from 63 to 500 microns at a height of less than 1.5 meters above ground level. Suspension is the longer-distance movement of particles with diameters smaller than 63 microns. Sandblasting is the technique of bombarding soil aggregates with saltating particles, producing aggregate fragmentation and the release of tiny particles that are then entrained (European Commission. Joint Research Centre., 2018).

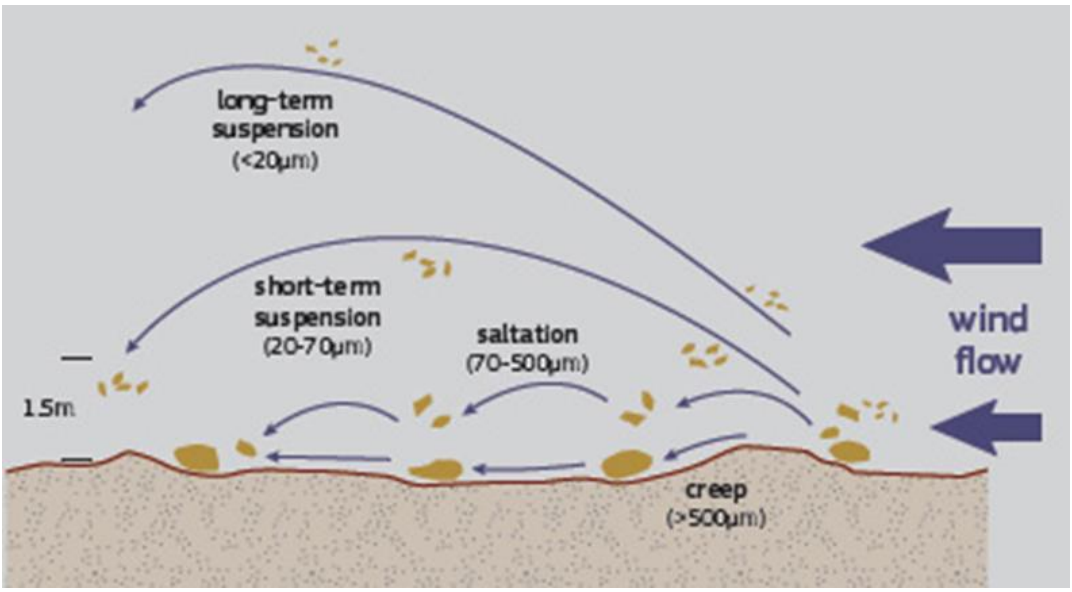

Figure 2.5: Dust particle transport processes (European Commission. Joint Research Centre., 2018)

In the Saharan region multiple dynamical mechanisms from micro to synoptic scales drive dust emissions and transport. After sunrise, powerful near-surface winds associated with northeasterly dry harmattan flow carry large volumes of dust, which is aided by vertical mixing of the nocturnal low-level jet (LLJ) momentum down to the surface (Lothon *et al.*, 2008). These winds meet the south-westerly moist West African Monsoon (WAM) flow near the surface at the intertropical discontinuity (ITD; Figure 2.6), which occurs over the Sahel from April to June (Lélé and Lamb, 2010). Furthermore, during the summer, mesoscale convective systems (MCSs) traveling through the Sahel often produce cold pools that spread rapidly over long distances and lift massive quantities of dust (Roberts and Knippertz, 2012). Moreover, African easterly waves (AEWs) propagating along the African easterly jet (AEJ) within the 10–20˚N latitudinal band (with average intervals of 3–5 days, detected at 700–850 hPa) are also important dynamical actors in the Saharan area (Fink and Reiner, 2003). AEWs, in conjunction with extratropical disturbances, significantly lead to the formation of a global dust hotspot over eastern Mauritania and northern Mali, by both emission and transport organization. Dust is often uplifted over this region by cold pools formed by convection connected with southerly moisture advection within AEW troughs (Knippertz and Todd, 2010). Dust is transported across the hotspot in the upper sections of the deep Saharan boundary layer by northerlies instead of AEW troughs (Chaboureau *et al.*, 2016). Over the summer, the ceiling of the boundary layer over the central Sahara can cross altitudes of up to 5–6 km (Cuesta *et al.*, 2020).

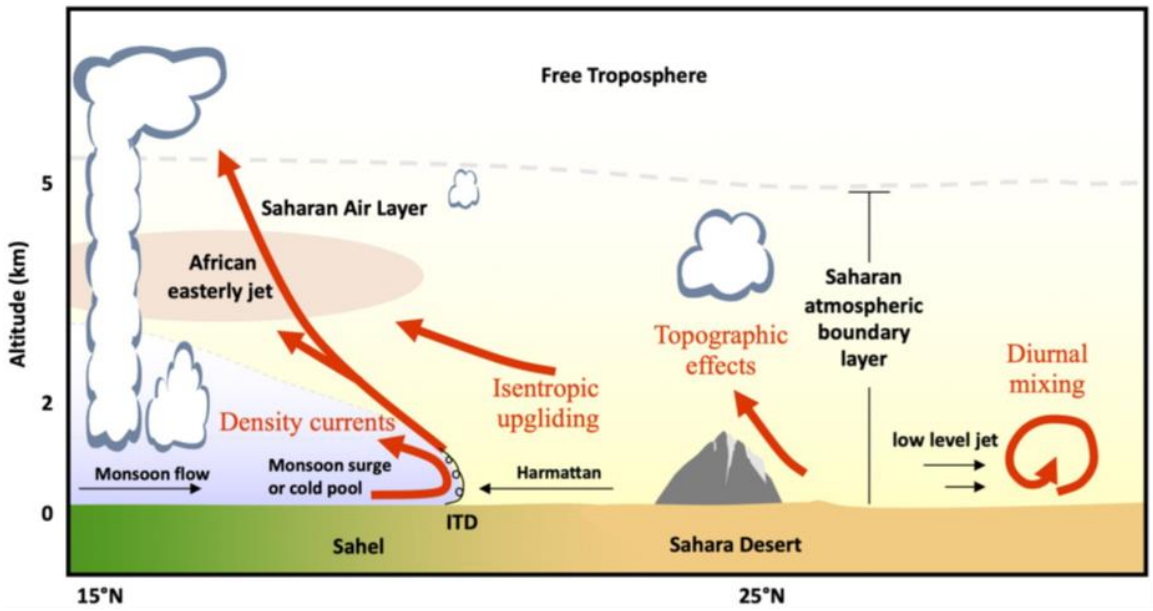

Figure 2.6: Schematic of mechanisms (red arrows) that control the vertical distribution of dust near the intertropical discontinuity (ITD). Shading (yellow and light blue) indicates air mass origins (from the Saharan boundary layer or the Gulf of Guinea, respectively) and temperature (Cuesta *et al.*, 2020)

While atmospheric conditions help to initiate the Saharan dust storms, when large quantities of Saharan dust particles are suspended in the atmosphere, they can have a noticeable impact on the lower part of the troposphere, which is referred to as the Planetary Boundary Layer (PBL) (Figure 2.7). The PBL is the layer of the Earth's atmosphere that is directly influenced by the Earth's surface, extending from the Earth's surface up to a few kilometres in altitude (Zhou *et al.*, 2021). It plays a crucial role in meteorology and atmospheric science because it is where most of our weather occurs and where pollutants and particulate matter are dispersed and mixed. Dust particles in the atmosphere can alter the energy balance of the PBL. During the day, these particles can absorb solar radiation, leading to warming and potentially causing the PBL height (PBLH) to rise as warm air tends to be less dense and more buoyant. Conversely, dust can emit longwave radiation

at night, cooling the PBLH and potentially lowering its height (Fluck and Raveh-Rubin, 2023). Also, dust aerosols can alter the vertical temperature profile in the PBL, which, in turn, can affect atmospheric stability. Depending on the specific conditions and properties of the dust aerosols, this can lead to changes in the PBLH (Zhang *et al.*, 2022). Moreover, Dust storms can influence the moisture content in the PBL by affecting local humidity levels. In some cases, moisture content changes can lead to shifts in the PBLH, especially in regions where moisture plays a significant role in atmospheric stability and convection (Francis *et al.*, 2022).

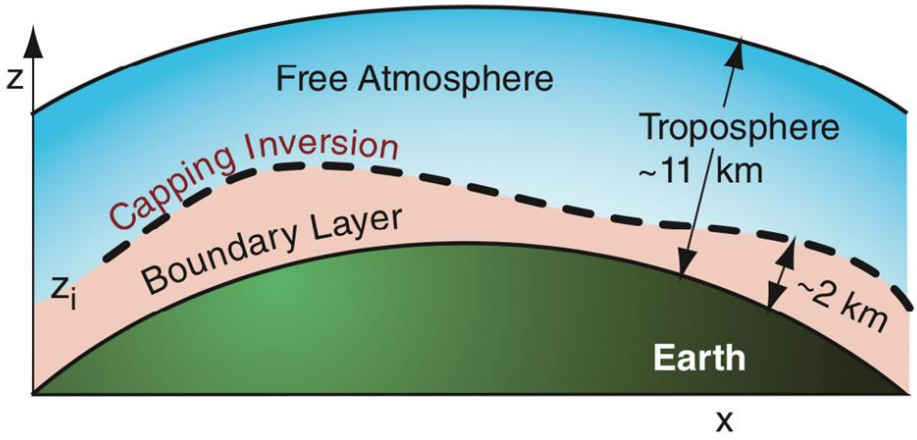

Figure 2.7: Planetary boundary layer illustration (Roland, 2018)

Saharan dust is frequently transported from its source areas along three major transport paths: westward across the North Atlantic Ocean to North America and South America; northward across the Mediterranean to southern Europe and occasionally as far north as Scandinavia; and eastward across the eastern Mediterranean to the Middle East as shown in Figure 2.8 (Goudie and Middleton, 2001).

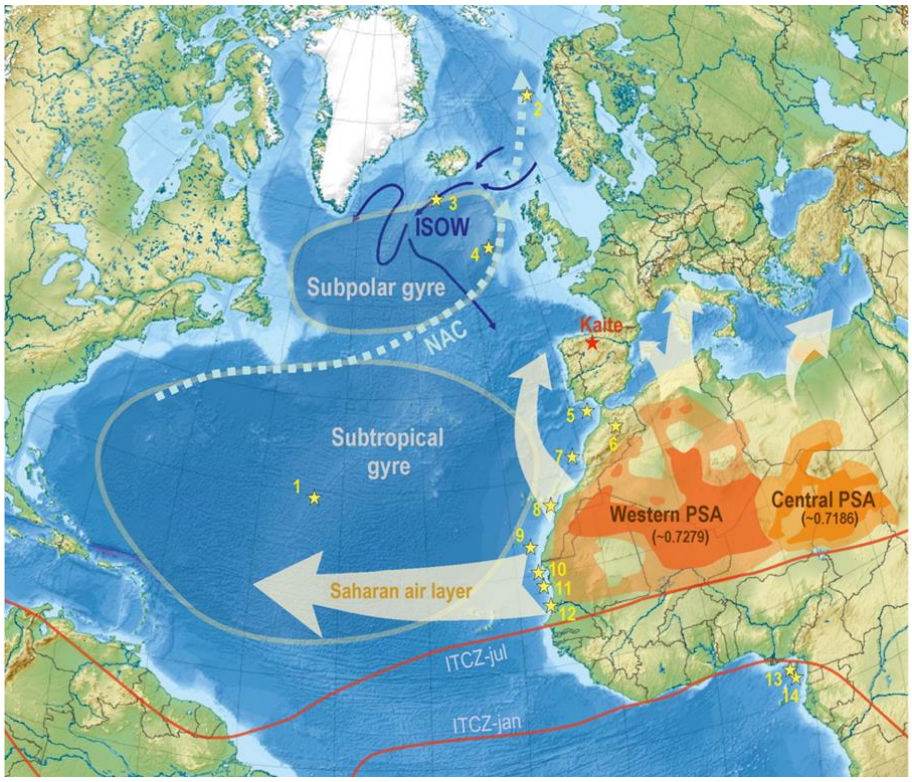

Figure 2.8: Major transport routes of Saharan dust (Cruz *et al.*, 2021)

2.2.1 Saharan dust transport to the Mediterranean, and European countries

The hundreds of thousands of tons of Saharan dust moved north have an impact on a wide range of environmental systems around the Mediterranean Sea. Moreover, Saharan dust effect the air quality levels in European countries during dust storm event periods, often the heavy Saharan dust transport result in the exceedance of the European Union standards concentrations of PM10 and PM2.5 (Pey *et al.*, 2013). Due to its geographical location, Spain suffers more from the Saharan Dust events. In his extensive study, Salvador *et al.* (2022) observed an increasing of the frequency and intensity of the Saharan dust events to the western Mediterranean in general. More in depth in other study, Salvador *et al.* (2019) analysed the pollution level in Madrid metropolitan area, and his results showed that intense Saharan dust occurrences raised the risk of daily death from PM10 exposure, and reduced the mixing layer height over Madrid.

The high PM concentrations during Saharan dust episodes was reported in many studies in the Mediterranean Basin and Western of Europe. Lyamani, Olmo and Alados-Arboledas (2005) highlighted that daily concentration of PM10 was 5 times higher during Saharan dust event days than in normal days in Spain. Concentrations of PM10 reached up to 250 µg/m^3 in Spain (Querol *et al.*, 2009), up to 470 µg/m^3 in Cyprus (Achilleos *et al.*, 2014), up to 156 µg/m^3 in Athens, Greece (Remoundaki *et al.*, 2011), up to 232 µg/m^3 in Portugal (Gomes, Esteves and Rente, 2022), and up to 135 and 113 µg/m^3 in Brixlegg Austria and Munich, Germany respectively (Bruckmann *et al.*, 2008).

In addition, The Moroccan Sahara and Atlas, as well as central Algeria, have been highlighted as key transit source areas for Western and southern Europe. These sources of transport have also been reported for the British Isles. A year of observation in Corsica, for instance, showed Twenty dust outbreaks from the following sources: Morocco/Western Algeria; eastern Algeria/Tunisia/Western Libya (Goudie and Middleton, 2001).

Saharan dust was found also to be transported to the Central European region. In extensive research done by Varga, Kovács and Újvári (2013), showed that the Carpathian Basin was a subject of 130 Saharan Dust episode from 1979 to 2011, with major period of dust transport occurred in the spring, with a subsequent maximum in the summer (in July and August), and dust events were also reasonably strong in February and October. Additionally, couple of source regions were identified such as Adrar and Tiris Zemmour region in Mauritania, Taoudénit in Mali, and Adrar and Tamanghasset in Algeria (Varga, Kovács and Újvári, 2013).

2.2.2 Saharan dust transport across the Atlantic Ocean

The transatlantic transport refers to the movement of Saharan dust westward over the Atlantic Ocean. This is the most extensive trajectory, accounting for 30-60% of total yearly Saharan dust, and it transports Saharan dust to the United States, South America, and Caribbean Islands (Touré, Konaré and Silué, 2012). The transport of the Saharan dust forms a layer of dry air and mineral dust that can have a thickness of about 3 km and its base starts from about 1.6 km above surface. This layer is referred as the Saharan Air Layer (SAL) (Luo and Han, 2021), and it can take place during early fall, late spring, and summer (Touré, Konaré and Silué, 2012).

The discovery of Saharan dust in the Caribbean Sea was first revealed in two papers that demonstrated the presence of large quantity of Saharan dust in measurement field in Barbados that was in 1965-1966 (Delany *et al.*, 1967; Parkin, Delany and Delany, 1967). After that, Prospero (1968), reported the first graph that describe the seasonal variability of the Saharan dust transport by combining data described in Delany *et al.* (1967) and data from Miami group, presenting evidence that would be later proof of the modulation impacts of climate on the dust transport after 50 years, even that there is many elements of the relationship between climate and African Saharan

dust formation and transport that are unknown and still a topic of debate (Evan *et al.*, 2016). Long-term aerosol investigations undertaken in the eastern Caribbean during the last 20 years have supplemented Barbados research (Prospero *et al.*, 2021). These investigations are being conducted in conjunction with research on aerosols, clouds, and precipitation in Puerto Rico (Valle-Di´az *et al.*, 2016). Furthermore, air quality monitoring on Martinique, Guadeloupe, and French Guiana, all departments of France, show an increase in PM concentrations whenever Saharan dust is present in the region's atmosphere, highlighting that a significant amount of nutrients transported in Saharan dust, which include phosphorus, seem to be crucial for maintaining soil fertility with in Amazon Basin (Prospero *et al.*, 2020).

During high Saharan dust season in 2007, high values of PM10 concentrations were recorded in the Caribbean basin, it reached 149 $\mu g/m^3$ in Martinique, 157 $\mu g/m^3$ in Guadeloupe, and 197 $\mu g/m^3$ in Puerto-Rico (Plocoste *et al.*, 2022).

Thus, the Saharan dust transport to different regions around the world, affect the air quality by affecting the PM10 and PM2.5 concentrations. However, it also provides important nutrients for marine and soil ecosystems.

2.2.3 Saharan dust transport across the eastern Mediterranean to the Middle East

The transport of the Saharan dust eastward was estimated to be 70 million tons of Saharan dust per year (Ganor and Mamane, 1982), where one-third of this volume reaches the eastern Mediterranean regions, which is about 2000 km from the Saharan source. In his research Cao *et al.* (2015) specified different dominant paths of Saharan dust transport to the Middle-East. There were three primary paths mentioned. The first path originates south of the Mediterranean Sea or the northern African coast and invariably strikes south of Syria or the Jordan-Saudi Arabia border. The second path is from north Africa, which normally goes via Egypt, north of the Red Sea, and blows southeast into Saudi Arabia.
The third path is also seen in depressions in northern Africa. It does, however, generally move south of Saudi Arabia through Sudan or the Red Sea. When facing monsoon storms from the Indian subcontinent, this course will occasionally curve upward to the east of Saudi Arabia or possibly Iran over the Persian Gulf. In addition, PM10 and PM2.5 always skyrocketed during dust event. Hussein *et al.* (2020) reported high concentrations of PM10 during May 2018–March 2019 dust episodes in Amman, Jordan. The 24-hour PM10 concentrations during Dust outbreaks ranged from 108 to 188 $\mu g/m^3$, a level that was about 3-6 times greater than the mean levels during clean circumstances (approximately 33 $\mu g/m^3$). In Kuwait also, in March – June 2021 dust event PM2.5 increased by 26.62% (Meo *et al.*, 2022). While in Ahvaz, Iran, for seven severe dust storms between 2009 and 2012, PM10 concentration was always high in a range between 137–553 $\mu g/m^3$ (Maleki *et al.*, 2022).

2.3 Estimating PM

Estimating particulate matter concentration from satellite data is a complicated procedure that includes data pre-processing, atmospheric adjustment, and retrieval of particulate matter concentration. Throughout the last decade, several methods for estimating ground PM concentrations have been developed. Although the effectiveness of satellite-based PM estimation models has continued to improve, there are still some limitations, such as missing values in satellite-based variables and PM ground measurements, spatiotemporal scale issues, and an unbalanced distribution of PM concentrations (Shin *et al.*, 2020). While satellite data gives vital information regarding particulate matter pollution, it does not provide a direct measurement of particulate matter concentration. Ground-based measurements are still required to confirm satellite data and offer more specific information regarding particulate matter pollution at the local level. Although, the density of current ground stations is restricted, and the majority of stations are

concentrated in metropolitan areas. Furthermore, regardless of how evenly or densely distributed they are, ground stations are point observations and hence fundamentally not continuous on the spatial domain. In contrary, satellite sensor systems have had the benefit of gathering data across wide areas on a frequent basis, thereby bypassing the spatial limits of ground measurements.

2.3.1 Satellite Remote Sensing

Remote sensing is the process of collecting information about objects or areas from a distance using sensors and other technologies. This technique allows scientists and researchers to gather data about the Earth's surface and atmosphere without physically being present. Remote sensing plays a crucial role in fields such as geography, geology, agriculture, meteorology, oceanography, and environmental management, as it provides a convenient and cost-effective way to acquire data about large areas in a short amount of time. By analysing this data, scientists can better understand patterns, trends, and changes in various parts of the world, and use this information to make informed decisions (Chapman and Gasparovic, 2022).

The history of remote sensing dates back to the mid-19th century, when scientists first began using aerial photography to study the Earth's surface from above. This early form of remote sensing used balloons and kites to carry cameras and other equipment into the air to capture images of the ground. In the 1920s and 1930s, the development of aircrafts allowed for more extensive aerial surveys, and by the 1950s, the first satellite-based remote sensing systems were being developed (Laffly, 2020). One of the earliest examples of satellite-based remote sensing was the Soviet Union's Sputnik 1, which was launched into orbit in 1957. Although Sputnik 1 was not designed specifically for remote sensing, its data provided valuable information about the Earth's environment. In the following years, a number of other satellite missions were launched, including the United States' Landsat program, which was established in 1972 and is still active today (Laffly, 2020). With the advent of digital technology, remote sensing has continued to evolve and improve. Today, remote sensing technology is used in a wide range of applications, including monitoring crop growth, detecting changes in land use and land cover, mapping natural disasters, and tracking the movements of wildlife, among others. The development of new sensors and advancements in computer processing power have also made it possible to process and analyse massive amounts of data in real-time, providing scientists with a wealth of information about the Earth's environment (Laffly, 2020).

One of the unique aspects of remote sensing space is the placement of satellites or platforms in orbit around the Earth. These orbits obey complicated principles and are classified into three categories based on target application and whether the entire or a portion of the earth sphere is to be monitored: geostationary orbit, medium Earth orbit, and low Earth orbit (Figure 2.9).

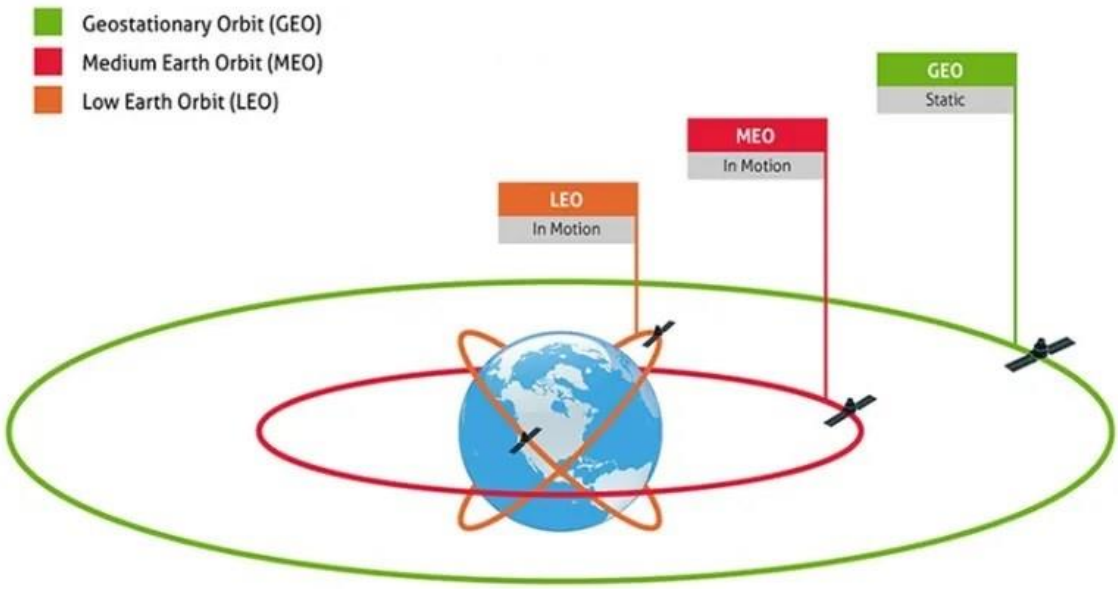

Figure 2.9: Types of Satellite orbits (‘Types of Satellite Orbits - mjginfologs.com’, 2020)

A geostationary orbit is one in which a satellite is located above the equator and seems to stay stationary in relation to a fixed point on the Earth's surface. This orbit is obtained by positioning a satellite at a height of roughly 36,000 kilometres above the Earth's surface and having an orbital period equal to the rotation period of the Earth. The satellite is always positioned above the same point on the Earth's surface in a geostationary orbit, allowing it to continuously monitor the same region. Geostationary orbits are therefore very valuable for applications like as weather forecasting, telecommunications, and remote sensing (Liang, Li and Wang, 2012). Geostationary satellites are launched into geostationary orbit by a rocket, which must launch the satellite with enough speed and direction to catch it into orbit. Once in geostationary orbit, the satellite is positioned with thrusters to stay stationary relative to the Earth's surface. Geostationary satellites have certain restrictions in addition to their stability and surveillance capabilities. Because they are so far away from the Earth's surface, their spatial resolution is weaker than that of satellites in lower orbits. Furthermore, because they are located above the equator, they have limited coverage of the Earth's poles, which limits their utility for some applications (Chapman and Gasparovic, 2022).

Satellites in the Medium Earth orbit are satellites that orbit at an intermediate height between Low Earth Orbit and Geostationary Orbit, between 10,000 to 20,000 kilometres above the Earth's surface. Medium Earth orbit satellites offer various benefits over low Earth orbit or geostationary orbit satellites. For example, because they are at a reasonable altitude, they provide a good mix of spatial resolution and global coverage, making them helpful for a variety of remote sensing applications (Lücking, Colombo and McInnes, 2012). Furthermore, because they are higher in altitude than low Earth orbit, they have a bigger field of vision and can cover larger regions of the Earth's surface in a single photograph. Navigation (e.g., GPS), earth observation, and telecommunications are some of the specialized applications of medium Earth orbit satellites. Medium Earth orbit satellites can also be used for environmental and climatic monitoring, as well as military surveillance and information collection (Löffler, Burgdorf and Klinkner, 2021). Nevertheless, medium Earth orbit spacecraft, have several drawbacks as compared to satellites in other orbits. For example, because they are higher in altitude than low Earth orbit, they have a longer orbital period, which means it takes them longer to return to the same location on the Earth's surface. Furthermore, because they are lower in altitude than geostationary orbit, they are exposed

to increased atmospheric drag, which can cause their orbits to degrade over time and necessitate more frequent orbit modifications (Kidder, 2015).

Low Earth orbit refers to a range of altitudes between 160 and 2,000 kilometres above the Earth's surface. Low Earth orbit satellites are placed very close to the Earth's surface, allowing them to cover a small region of the Earth's surface in a single photograph with great spatial resolution. The great spatial resolution of Low Earth orbit satellites enables for comprehensive studies of the Earth's surface, which is one of their key advantages. As a result, they may be used for earth observation, remote sensing, and scientific study (Rossi, Petit and McKnight, 2020). Furthermore, because they operate at a lower height than higher-altitude satellites, they can see the Earth's surface more often, enabling for real-time monitoring of occurrences such as natural catastrophes and environmental changes. Earth observation, weather forecasting, and remote sensing are some of the specialized uses of low Earth orbit satellites. Low Earth orbit satellites may also be used for military reconnaissance and information collection, as well as tracking ship and aircraft whereabouts. However, compared to satellites in higher orbits, low Earth orbit spacecraft have several drawbacks. Because they are at a lower altitude, they have a restricted field of view and can only cover a tiny portion of the Earth's surface in a single photograph. Furthermore, because to their proximity to the Earth's surface, they are exposed to increased air drag, which might cause their orbits to deteriorate over time and necessitate more frequent orbit modifications (Montenbruck and Gill, 2000).

The National Aeronautics and Space Administration (NASA) has launched several satellites for atmospheric observations, each with a unique set of instruments and capabilities (Figure 2.10). There are currently many satellites which products are committed to be accessible by NASA from public sites under a "free and open" data policy than ever before. NASA have a broad range of Earth observing data, with new information pertinent to understanding widespread changes in Earth conditions caused by warmer temperatures, global changes in precipitation patterns, continued conversion of earth resources to human uses, and invasive species and biodiversity loss, as well as a long list of natural disasters (hurricanes, wildfires, earthquakes, volcanic eruptions, tsunamis, floods, aerosols, etc.) (Ustin and Middleton, 2021).

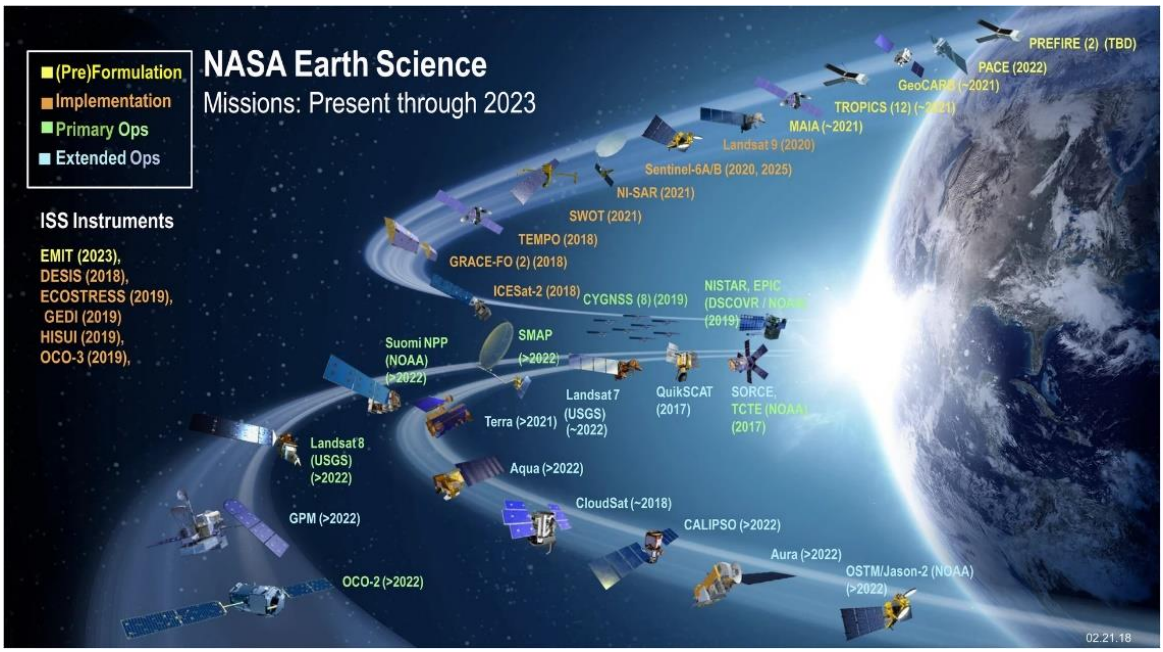

Figure 2.10: NASA’s active Earth science satellite missions in 2019 (Ustin and Middleton, 2021)

The amount of aerosol in the atmosphere, as measured by AOD, characterizes the degree to which solar radiation is dispersed and absorbed by particles (Shin *et al.*, 2020). In other words, dust, smoke, and pollution in the atmosphere can block sunlight by absorbing or scattering light. AOD informs us of the amount of direct sunlight that these aerosol particles keep from reaching the ground. It is an arbitrary number with no dimensions that describes how much aerosol is present in the vertical column of atmosphere. A value of 0.01 indicates an environment that is exceptionally clean, while a value of 0.4 indicates an environment that is extremely hazy (Zhang *et al.*, 2020). There are two basic methods for measuring the AOD of the atmosphere. One method is to employ ground sun photometers, while the other relies on sensors onboard satellite platforms. Ground-based AOD observations give real-time AOD data, which aids in the calibration of satellite-borne AOD devices and improves the prediction performance of remote sensing retrieval algorithms. The AERosol RObotic NETwork (AERONET) is a significant ground-based AOD monitoring network that has been operational for over 25 years and consists of over 500 locations throughout the world. This network effectively confirms satellite-based AOD retrievals, but its operation is costly and insufficient due to its limited spatial coverage, particularly in less populated regions (Wei *et al.*, 2020). In contrast, satellite remote sensing can measure AOD with wide-area spatial coverage on a global or regional scale, including inaccessible areas. As a result, ground-based AOD observations and satellite-derived AOD databases complement each other. Because of the availability of numerous Earth-observing satellite sensors, satellite-derived AOD products were accessible beginning in the 1980s (Wei *et al.*, 2020).

Technical improvements have enhanced AOD retrieval techniques for retrieving data from the many satellites and research equipment involved. Various papers (Mishchenko and Geogdzhayev, 2007; Kim *et al.*, 2014; de Leeuw *et al.*, 2015; Luo *et al.*, 2015) discuss the advantages and disadvantages of various retrieval algorithms. Satellite signals at the top of the atmosphere level are derived from reflected and scattered reflectance from the ground surface and atmospheric particles. The main difficulty in calculating AOD is distinguishing the contribution of surface scattering as well as atmospheric signals in the top of the atmosphere reflectance accurately (Wei *et al.*, 2020). Because the reflectance of red and longer wavelengths is essentially negligible over open ocean, this problem is mostly restricted to retrieving AOD over the land surface. In these wavelengths, ignoring surface reflectance over the ocean may result in a decent approximation of AOD (Levy, Remer and Dubovik, 2007).

Geostationary satellite sensor systems particularly, give data with high temporal precision (minutes), that is ideal for monitoring air quality. Estimation of ground PM concentrations via satellite data has received a lot of attention, especially when it comes to aerosol products like aerosol optical depth (AOD). Among a variety of satellite sensor data, AOD data from the Moderate Resolution Imaging Spectroradiometer (MODIS) have dominated utilization in the literature. Studies based on MODIS-based AOD data obtained by alternative AOD retrieval techniques, such as the Multi-Angle Implementation of Atmospheric Correction algorithm and the Simplified Aerosol Retrieval Algorithm, account big proportion of published papers in the domain. This is due to MODIS's lengthy operational term (MODIS Terra: 1999-present, MODIS Aqua: 2002-present), as well as its geographical and temporal resolutions, which are suitable for regional, continental, and even global studies. However, MODIS's planned lifetime has gone, and additional satellite data, such as Visible Infrared Imaging Radiometer Suite (VIIRS), may be required as a replacement in the future (Shin *et al.*, 2020).

2.3.2 Machine Learning methods

Machine learning techniques are a subset of algorithms that enable computers to recognize patterns in data and predict outcomes based on those patterns. These approaches are rapidly being used to solve a wide range of remote sensing challenges, including the estimate of particulate matter

concentrations using satellite data (Dobrea *et al.*, 2020). The application of machine learning techniques in remote sensing implies training a model on a big database of satellite data and ground-based particulate matter measurements. The purpose of this training is to understand the association among satellite data and particulate matter concentration to ensure that the model can forecast particulate matter concentrations accurately based on newer satellite data (Géron, 2022).

Machine learning algorithms have been widely used to forecast air quality, capturing the interest of scientists worldwide. A total of 2090 articles were published in the English language between 1992 and 2021, with 658 articles published between 1992 and 2016 and 1432 articles published between 2017 and 2021, indicating a significant increase in publications that is 2.17 times higher between 2017 and 2021 than 1992-2016. This suggests that machine learning-based prediction tools are growing more popular as a result of the availability of a diverse set of machine learning packages. In terms of the number of publications, China and the United States emerged as the top two countries. China generated the most articles (687), then the United States (264 articles), India (162 articles), and Spain (162 articles) (Mehmood *et al.*, 2022).

In early research papers, they detailed various approaches for using satellite products to get surface PM2.5. Different satellite used in research articles (Gupta and Christopher, 2009b). In general, columnar satellite-derived AOD values at 0.55 µm are connected to surface PM2.5 mass observations. The satellite AOD data are then converted to PM2.5 mass concentration and ultimately to air quality indexes using a linear regression method. It is vital to emphasize that AOD is an optical measure of the column loading of atmospheric aerosols and it is unitless, whereas PM2.5 is the mass concentration of particles smaller than 2.5 µm in aerodynamic diameter measured at the surface in units of $\mu g/m^3$. By understanding the aerosol microphysical and optical characteristics, as well as the height of the aerosol layer in the atmosphere, satellite-measured AOD may be empirically translated into PM2.5 mass (Gupta and Christopher, 2009b). Though, many studies, have found that the PM2.5-AOD connection cannot be used alone to determine surface level PM2.5 because the vertical distribution of aerosols and other meteorological characteristics such as humidity and temperature may also be essential (Liu *et al.*, 2005).

Van Donkelaar, Martin and Park, (2006) calculated the ground-level concentration of PM2.5 using space-based data from MODIS and the Multi-angle Imaging Spectroradiometer (MISR) satellite sensors, as well as supplementary information from a global chemical transport model (GEOS-CHEM), from January 2001 to October 2002. Surface data from the National Air Pollution Surveillance (NAPS) network in Canada and the Air Quality System in the United States were compared to remote-sensed PM2.5 (AQS). Their results showed that when calculated from MODIS ($r = 0.69$), and MISR ($r = 0.58$), the geographic variance in annual mean PM2.5 showed significant agreement with surface data, where r is the linear correlation coefficient between estimated and observed PM2.5. For both MODIS and MISR, the daily variation in remote-sensed PM2.5 was more consistent with surface observations in eastern North America ($r = 0.5$-0.8) than in western North America ($r = 0$-0.35). The conclusion was that satellite observations of AOD have the ability to provide a distinct overview of worldwide ground PM2.5 concentrations when combined with additional information from a chemical transport model on the link of AOD and PM2.5, and that the quality of aerosol remote sensing and model modelling of aerosol characteristics will determine the development of this capacity (Van Donkelaar, Martin and Park, 2006).

Gupta and Christopher, (2009b), did an estimation of PM2.5 concentration over the southeastern United States using MODIS AOD data, and coupling it with meteorological factors such as wind speed, temperature, relative humidity and boundary layer heights by using multiple regression approach, with a goal to evaluate if the meteorological factors will improve the estimation of PM2.5 based on AOD data. The results show that incorporating meteorological data via multiple

regression methods improves correlation values up to thrice when comparing to two variant regression equations (AOD and PM2.5). When temperature and boundary layer heights are added to the AOD-PM2.5 connection, the root-mean-square error improves by 20-50%. Summer and well-mixed boundary layer conditions had the best agreement between AOD and PM2.5. even though, their technique demonstrates that simple statistical models, rather than complicated physical and numerical models, may be utilized to predict PM2.5 with an average error of 34% for hourly and 24% for average daily mass concentrations. In addition, as part two for their study Gupta and Christopher, (2009a) instead of using multiple regression approach, they used an artificial neural network (ANN) framework to minimize the uncertainties of surface PM2.5 estimate using satellite data. Comparing the outcome of the study to regression coefficients acquired using simple correlation (R = 0.60) or multiple regression (R = 0.68) procedures, the ANN obtains hourly PM2.5 data with R = 0.74 when compared to observational data. The ANN approaches yield a correlation of R = 0.78 for calculating daily mean PM2.5.

As machine learning continued to develop, more and more papers were published to estimate both PM10 and PM2.5 using different machine learning algorithms such as Multiple linear Regression (MLR), ANN, Random Forest (RF), and Support Vector Regression (SVR).

In general, MLR accuracy was relatively poor in numerous studies, with value of R^2 less than 0.25, whilst the best model in each research had R^2 larger than 0.49 (Shin *et al.*, 2020).

In PM estimation research, SVR a representative technique based on the kernel method, has been compared to various methodologies. Li *et al.* (2018) compared four approaches (decision tree, orthogonal regression, SVR, and RF). The Correlation coefficients of the SVR models for MODIS Terra and Aqua AOD were 0.72 and 0.78, respectively, ranking them second among the four models behind RF (0.77 and 0.85 for MODIS Terra and Aqua AOD respectively). Yeganeh *et al.* (2017) utilized SVR and several types of ANN algorithms to predict monthly average PM2.5 in Australia's South-east Queensland area between 2006 to 2011, based on AOD and detailed meteorological and geographical data. The findings showed that the R^2 of one of the ANN (0.81) was better than that of the SVR (0.67), and the conclusion was that employing ANN algorithms in association with satellite remote sensing data, geographical information, and meteorological data improved PM2.5 estimate.

Numerous research has looked into using the Random Forest technique to estimate PM2.5 ground surface concentration. In terms of the accuracy and reliability of the RF model predictions, studies have shown encouraging findings. Hu *et al.* (2017) used the RF algorithm to predict daily PM2.5 concentrations over the United States in 2011 by combining meteorological fields, land use factors, and AOD data. The study found that the RF model achieved good performance in terms of predicting PM2.5 concentrations, with an R^2 of 0.8. An RF model was also employed in Chen *et al.* (2018) to estimate the PM2.5 concentration in China between 2005 and 2016. A mix of meteorological, AOD, and ground-level PM2.5 datasets were used to train the model, and the results demonstrated that the RF model was very accurate in forecasting PM2.5 concentration with an R^2 of 0.83 for daily estimated PM2.5 and 0.86 for monthly and annual PM2.5 average concentrations. Ghahremanloo *et al.* (2021) compared estimations of daily 1 km PM2.5 concentrations in Texas between 2014 and 2018 of RF model to others models like MLR and the mixed effect model. In the study the machine learning models were trained using MODIS AOD data coupled with meteorological data and MERRA-2 data such as dust, sea salt, OC, BC, and SO_4. the findings demonstrated that RF had a strong capacity and high accuracy in estimating PM2.5 concentrations across the study region as compared to other models with a correlation coefficient in the range of 0.83–0.90 and a mean absolute bias between 1.47 and 1.77 $\mu g/m^3$.

2.4 Summary of the literature review

The study of the literature has revealed that PM particles originate from various sources depending on the region, and the study of PM can be done from multiple perspectives.

The scientific literature on Saharan dust storms, particle matter (PM) concentration monitoring, and integrating satellite data with machine learning provides an in-depth look at air quality dynamics. Several studies have examined the strength of Saharan dust storms and their impact on PM concentrations, revealing a significant rise in coarse and fine PM levels during those events, affecting air quality and human health. Many conclusions go that the Saharan dust storms' intensity and timing are changing due to climate change effects, as do the effects of Saharan dust on the PM concentrations. Additionally, there is growing interest in using low-cost PM sensors to monitor PM concentrations to study PM in more depth, providing cost-effective alternatives for extensive monitoring. However, accuracy, calibration, and reliability remain challenges.

Satellite remote sensing has developed as a significant technique for monitoring PM concentrations across vast geographical areas, applying algorithms and models to predict ground-level PM concentrations based on satellite-derived data and other atmospheric characteristics. Furthermore, machine learning methods are increasingly being used to improve the precision of PM concentration estimation, particularly when satellite data is used. These methods entail training models on ground-based monitoring data and satellite observations to capture complicated interactions and increase the accuracy of PM level estimates. Despite progress, there are still issues in precisely estimating the impact of Saharan dust storms on PM concentrations, assuring the dependability of low-cost sensors, and refining machine learning algorithms. Also, numerous research papers have estimated ground surface PM concentrations using satellite data in many regions of the world, but no study has been found for Hungary.

Ongoing research aims to address these issues by integrating multi-sensor data, improving sensor calibration, and developing sophisticated machine-learning models for robust PM concentration predictions. These all contribute to a better understanding of air quality dynamics and provide valuable tools for monitoring and managing particulate matter pollution.

3 MATERIALS AND METHODS

In this chapter I describe the tools and datasets I used, as well as the methods and relationships used in data processing.

This chapter is divided into four parts. First part contains experimental section, which is two subchapters, a small-scale experiment of PM10 dispersion around obstacles, and the effect of small hills on PM10 and PM2.5 concentrations in short range. Second is the part that deals with Saharan dust transport, divided into three subchapters, the Saharan dust event of June 2020, the identification and evaluation of the Saharan dust storm events to Hungary between 2018 and 2022, and the case study of the Saharan dust effects on PM10 and PM2.5 concentrations in Budapest in March 2022. The third section is the estimation of PM concentrations, and contains two subchapters, is the Evaluation of PM surface concentrations simulated by Version 5.12.4 of NASA's MERRA-2 Aerosol Reanalysis over Hungary in the period between 2019 and 2021 using two approaches to estimate PM ground-level concentrations using surface, satellite, and meteorological data based on machine learning algorithms, and, the calibration of CAMS PM2.5 data over Hungary using machine learning. The fourth part is subchapter that describe the common data and statistics that are used throughout the study.

3.1 PM dispersion experiments

3.1.1 Small scale experiments of PM10 dispersion around obstacles

Small scale experiments were conducted in order to investigate the effects of obstacles heights and distance from the source in the PM10 concentration. The goal was to understand the changing of the PM10 concentration around obstacles in simple environment. The experiments were done in isolated laboratory room on built table. The table had 3 PM10 sensors with 50 cm distance between each sensor. The room temperature was stable during the experiments (25 ±1˚C), the same was for the Relative Humidity (RH) (50% ±3).

3.1.1.1 Experiment set up

The experiments were done with two different wind speed (air flow speed of 2.9 and 1 m/s measured by Schiltknecht MiniAir64 vane anemometer) provided by two different ventilators. The use of the ventilators is to make sure that the PM plume will follow the wind direction toward the sensors and to avoid the spreading of plume around the room. As mentioned, three sensors were used, sensor A, B and C as shown in Figure 3.1, sensor C placed near the source, sensor B in the middle and sensor A is 1 meter away from the source. The obstacle was placed at three different distances between sensors A and B, with changing of the obstacle height (12, 24 and 36 centimetres). The PM sensors used are NOVA PM sensors (SDS011) that use principle of laser scattering to get the particle concentration in the air, with a digital output and built-in fan that is stable and reliable.

The incense sticks were used as a source of PM10 plumes, due to the number of particles emitted from incense smoke in a short time. There were many research studies that investigated the effect of the use of incense sticks on PM10 concentrations. Numerous studies indicate that the smoke from burning incense contains particulate matter, gas products, and other organic compounds that can increase PM concentrations, CO, NOx, and SO_2 in the air (Jetter *et al.*, 2002; Ji *et al.*, 2010). Also, incense burning was found to increase PM2.5 concentrations by up to 120% (Tran *et al.*, 2021).

Each experiment took 15-20 minutes, by burning one incense stick with fixed wind speed, obstacle distance from the source and obstacle height. The total number of variations (experiments) was 18.

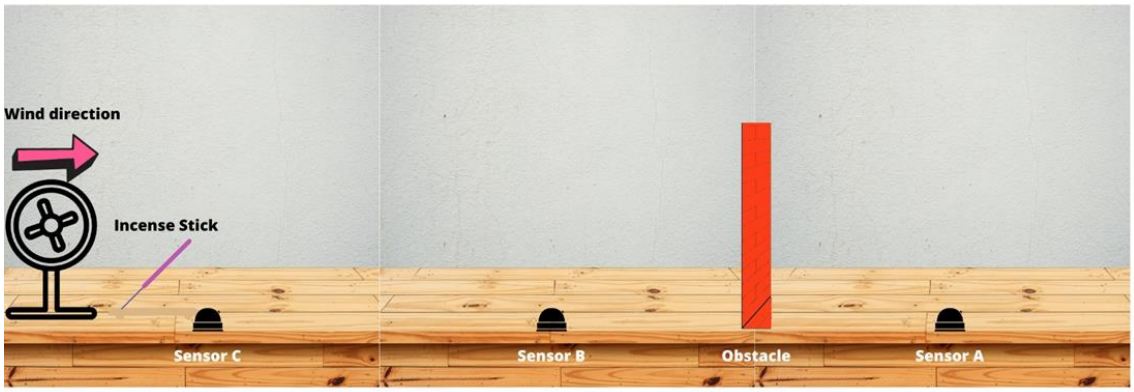

Figure 3.1. Experimental set up

3.1.1.2 Data analysis

Measurements were registered continuously in a programmed excel sheet during each experiment for every 30 seconds. The results present the average PM10 concentration in each test and presented in graphs depending of the obstacle height and distance from the source for the three sensors.

Also, we used the Multiple Linear Regression (MLR) method to calculate the PM10 concentration in sensor A ($PM10_A$) depending on the obstacle height (OH) and distance from the source (OD), PM10 concentration measured near the source (Sensor C, $PM10_C$), and wind speed (Ws).

3.1.2 Effect of small hills on PM10 and PM2.5 concentrations in short range

This part of my study aims to discover how small hills affect the PM10 and PM2.5 concentrations in short range with different wind speeds.

Figure 3.2 shows the experimental setting environment, where the PM sensors are hung on steel infrastructure where sensor 1 (S1) is close to the source (smoke machine), and sensor 2 (S2) is a sensor placed in the middle of the slope. Sensor 3 (S3) is at the top of the hill. The sensors used are low-cost sensors that are calibrated and used in other cities in Hungary to monitor the PM concentrations. The sensors were developed for a project called HUNGAIRY, which is a project that aim to improve air quality at 8 Hungarian regions through the implementation of air quality plan measures, where until this moment PM sensor (LIFE IP HUNGAIRY project sensor) is used in 60 PM monitoring stations in Miskolc and 20 PM monitoring stations in Kaposvár, and the network of the PM monitoring stations will be expended to other Hungarian cities (LIFE IP HungAIRy, 2019). The sensor is based on low-cost, laser scattering PM sensor (Plantower PMS7003), and an auxiliary sensor (Bosch BME680) for measuring humidity, temperature and pressure coupled with a Raspberry Pi 3 single-board computer to collect and store measurements (Báthory *et al.*, 2022). The smoke machine (Haze machine hs-600) is a machine used in concerts and festivals to generate smoke, and the wind machine is controlled via a variable frequency drive (VDF). In contrast, wind speed and directions were used to ensure that no wind was disturbing the experiments and that the wind was going in the right direction. In addition, the sensors register the PM2.5 and PM10 concentrations each minute.

Three cases were adopted in this study. Case 1 is where the ground is almost flat, case 2 is where there is a small hill with an elevation of 0.8 m (shown in Figure 3.2), and case 3 is where a higher elevation is 1 m. The place where the experiments took place was the backyard of a laboratory.

The experiments were done many times, and each time, the smoke machine was on for one hour because, after one hour, the performance of the smoke machine was not stable. The first 10 minutes are without any wind, and then every 10 minutes of wind, the frequency is increased via VDF until

we have the maximum frequency possible (here, five frequencies were used). Each wind frequency corresponds to wind speed measured by the wind speed sensor (presented in Table 1).

Also, we used the Multiple Linear Regression (MLR) method to calculate the PM10 concentration in sensor 3 ($PM10_{S3}$ in $\mu g/m^3$) based on the PM10 concentration near source (concentration registered by Sensor 1, $PM10_{S1}$), PM10 concentration at the bottom of the hill (concentration registered by Sensor 2, $PM10_{S2}$), the wind speed (Ws in m/s), and the height of the hill (H in m).

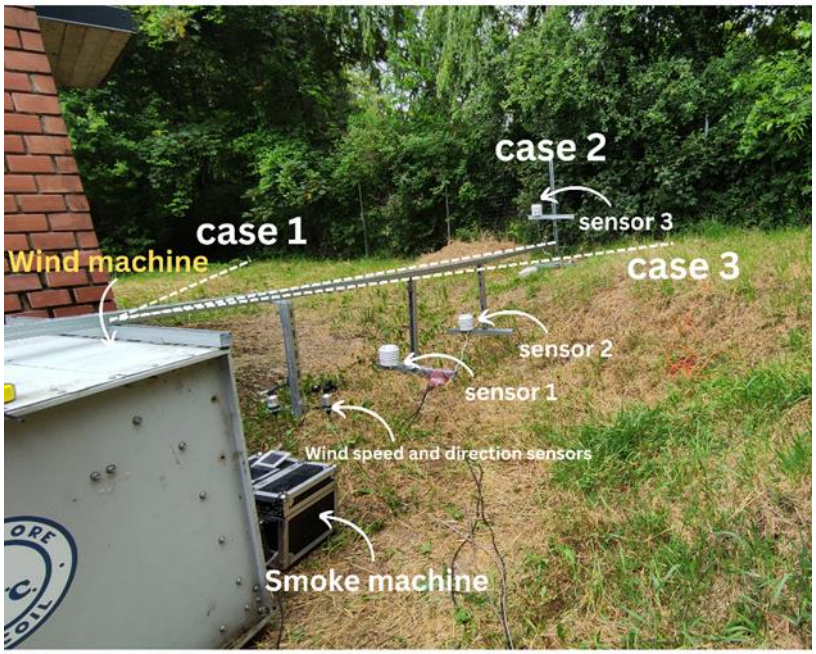

Figure 3.2. Experimental setting environment where dashed lines represent the placement of the metal structure in each of the three cases

Table 1: Wind speed depending on VDF output frequency

VDF output frequency (Hz)	wind speed (m/s)
10	0.7
20	2.4
30	3.7
40	5.1
50	6.1

3.2 Saharan Dust storm transport

3.2.1 Dust Storm simulation over the Sahara Desert (Moroccan and Mauritanian regions) using HYSPLIT

June 2020 was a month where a breaking record dust storm occurred over the Sahara and transported toward the Americas. According to Francis *et al.* (2020), the dust clouds that were generated in this event registered the highest record of Aerosol Optical Depth (AOD). The dust emission was continuous for 4 days, and uplifted to 5-6 km above the ground surface, and transported across the tropical Atlantic oceans by the powerful mid-atmospheric winds that had a speed higher than 20 m/s (Francis *et al.*, 2020).

For (Francis *et al.*, 2020) the primary objective was to determine the processes responsible for the lifting and transport of dust during the dust event, as well as their relationship to large-scale

circulation, and focus on the characteristics of the atmospheric mechanisms that led to massive transport of the Saharan dust. While in this research the central goal is to locate the most active regions in Western Sahara during that event, and the contribution of those regions in increasing the level of PM10 concentration in some regions that are far away from the source place like the US coastal part of the Gulf Mexico and the Martinique islands.

The dust clouds generated covered a huge space as shown in the true color images of MODIS-Aqua satellite on the 14, 15, 16, 17, 18, and 19 June 2020 (Figure 3.3), where the dust in yellow color is spreading from the Western Saharan region to the Atlantic Ocean.

3.2.1.1 HYSPLIT model description

The Hybrid Single-Particle Lagrangian Integrated Trajectory model (HYSPLIT) is a software developed by the Air Resources Laboratory (ARL) of the National Oceanic and Atmospheric Administration (NOAA) of USA (Draxler and Hess, 1998). The model is a comprehensive system for simulating basic air parcel trajectories as well as complicated transport, dispersion, chemical transformation, and deposition scenarios. The model calculation method is a hybrid of the Lagrangian approach, which uses a moving frame of reference to calculate advection and diffusion as trajectories or air parcels move away from their initial location, and the Eulerian methodology, which uses a fixed three-dimensional grid as a frame of reference to compute pollutant air concentrations. Over more than 30 years, the HYSPLIT model has developed from predicting simplistic single trajectories based on radiosonde measurements to a system that accounts for numerous interacting pollutants carried, dispersed, and deposited on local to global scales. In addition, HYSPLIT was used to assess the consequences of the accidental release of nuclear material into the atmosphere from the Fukushima Daiichi nuclear power plant after an earthquake and tsunami in March 2011. NOAA's interest since the middle of the last century at the latest, and modelling the movement of smoke from large wildfires has been an ongoing development activity at ARL since 1998. Today, in addition to the United States, smoke forecasts for Alaska and Hawaii are conducted daily to provide to air quality forecasters and the public information on fine particles (PM2.5) in the air (http: //airquality.weather.gov/) (Stein *et al.*, 2015).

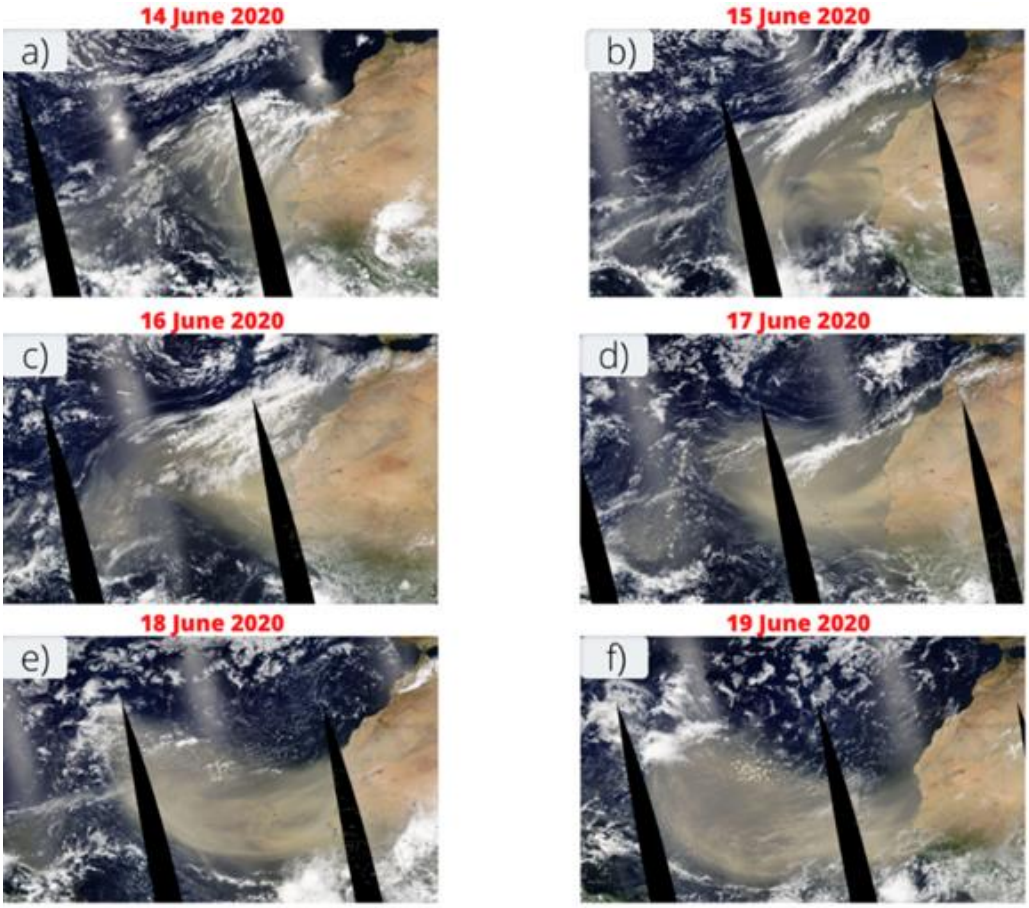

Figure 3.3. MODIS-Aqua true color images on the (a) 14th, (b) 15th, (c) 16th, (d) 17th, (e) 18th, and (f) 19th of June 2020 over western Africa and the northern tropical Atlantic Ocean. The white color represents the Clouds and the yellow the dust.

Dust storm Model

HYSPLIT dust storm model is a model for the emission of PM10 dust that has been built using the theory of a surface-roughness-dependent threshold friction velocity (Draxler *et al.*, 2001). When the local wind velocity exceeds the threshold velocity for the soil properties of that emission cell, a dust emission rate is computed from that model grid cell. The predominant mechanism for PM10 emission is "sandblasting," in which larger particles that cannot go airborne bounce along the surface (saltation), allowing additional smaller particles to become airborne (Draxler, Ginoux and Stein, 2010). This emission module makes use of HYSPLIT's 1° land-use file, assuming that a "desert" land-use grid cell corresponds to the roughness identification class "active sand sheet." Only on dry days and when the friction velocity exceeds the threshold value (0.28 m/s for an active sand sheet) do dust emissions occur. Once the emission strength is determined, the model emits Lagrangian particles with a mass calculated by multiplying the PM flow by the 1° area corresponding to a desert category in HYSPLIT's land-use file. These Lagrangian particles are distributed and moved forward in time in response to NOAA's GFS model's meteorological fields with a horizontal resolution of 1°. A more specific description of particle dispersion and transport can be found in (Draxler *et al.*, 2001; Escudero *et al.*, 2006).

The meteorological data fields needed for the model can be accessed from the National Climatic Data Centre (NCDC) website which is NOAA's National Centers for Environmental Information (NCEI) that provides public access to remarkable archives for environmental data on Earth. In this study, we used the GDAS (Global Data Assimilation System) meteorological data (GDAS1) with

a horizontal resolution of 1˚×1˚ corresponding to approximately 100 km x 100 km and 23 vertical layers. GDAS1 is chosen to match the resolution of the HYSPLIT land-use file resolution. GDAS is a system used by the Global Forecast System (GFS) model to insert observations into a gridded model space to begin or initialize, weather predictions using observed data. Surface observations, balloon data, wind profiler data, airplane reports, radar observations, and satellite observations are all added to a gridded, 3-D model space by GDAS. GDAS data are provided as both GDAS input observations and GDAS gridded output fields. The GFS model can be started using gridded GDAS output data. Input data are accessible in a number of data formats due to the varying nature of the assimilated data types, notably Binary Universal Form for the Representation of Meteorological Data (BUFR) and Institute of Electrical and Electronics Engineers (IEEE) binary. World Meteorological Organization (WMO) Gridded Binary is the GDAS output (GRIB) (Kleist *et al.*, 2009). The GDAS dataset covers the entire globe and is freely available.

In the dust storm model, the study domain is defined from 15.0N -18.0E to 32.0N -05.0E (Domain covered with stars in Figure 3.4) which covers the Western Sahara of Morocco Mauritania, and a small part from Algeria. While the PM10 concentrations are averaged over every 12h. The dust simulation Started on the 14th of June 2020 at 00UTC until the 19th of June 2020 at 00UTC. HYSPLIT dust storm modelling was set for 0.5°x0.75° resolution for desert dust sources, with a total of 10 million particles or puffs released during one release cycle and a maximum of 5 million particles permitted to be carried at any time during the simulation. The release mode is sampled using three-dimensional particles in both horizontal and vertical orientations.

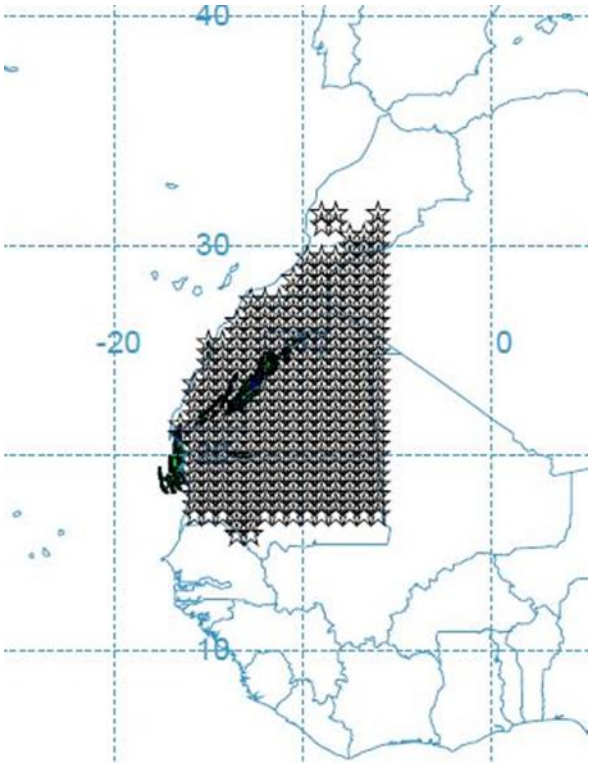

Figure 3.4. Map of the study domain

Trajectory cluster analysis

Forward and Backward trajectory analysis are reliable methods to identify the long-range transport patterns with the use of archived meteorological data (Baker, 2010). However, considering the benefits of the trajectory model, individual trajectories are subject to errors due to the precision and quality of the meteorological data, as well as the simplifying assumptions employed in the trajectory model, which ultimately limits their utility. This problem was solved by computing a

large number of trajectories and then subjecting them to cluster analysis. The large number of trajectories computed in HYSPLIT trajectory cluster analysis refers to the number of individual trajectories generated and then subjected to cluster analysis. The benefits of computing a large number of trajectories include minimising the effects of individual trajectory errors, providing a more comprehensive picture of the atmospheric conditions, and identifying rare or unusual events that a smaller number of trajectories may not capture (Baker, 2010). The exact number of trajectories computed will depend on the specific analysis being conducted and the available computing resources. The differences among these trajectories are determined by calculating the distance between clusters, with smaller distances indicating higher similarity. The clustering computation minimises the differences between trajectories within a cluster while maximising the differences between clusters. Trajectories are combined until the total variance of the individual trajectories about their cluster mean starts to increase. This occurs when disparate clusters are combined. The clustering computation is described in more detail in the literature (Shaw and Gopalan, 2014).

HYSPLIT forward trajectory cluster analysis was performed for the regions that are considered the most active sources of dusts and particles in the region of study as long as some surrounding regions. The list of the regions is presented in Table 2 with names, latitudes, longitudes, and time periods.

Table 2. Cluster analysis location lists

Location	**Latitude**	**Longitude**	**Simulation period**
Dakhla, Morocco	23.8	-15.6	10-30 June 2020
Bir Anzarane, Morocco	23.88	-14.53	10-30 June 2020
Oum Dreyga, Morocco	24.1	-13.25	10-30 June 2020
Aousserd, Morocco	22.5	-14.3	10-30 June 2020
Nouakchott, Mauritania	18.09	-15.95	10-30 June 2020
Atar, Mauritania	20.5	-13.05	10-30 June 2020
Tichit, Mauritania	18.45	-9.5	10-30 June 2020
Toumbouctou cercle, Mali	20.0	-3.0	10-30 June 2020
Bordj Badji Mokhtar, Algeria	22.62	0.12	10-30 June 2020
Tamanrasset, Algeria	24.37	4.32	10-30 June 2020

3.2.1.2 Satellite Observations

Satellites are increasingly being utilized to collect data on aerosol features such as aerosol optical depth (AOD), the columnar concentration of particles, and particle sizes, taking advantage of technological and scientific advances over the previous years. There are various Earth Observing satellite instruments that developed many aerosols remote sensing algorithms for the retrieval of the AOD. One of those instruments is the Moderate Resolution Imaging Spectroradiometer (MODIS). The MODIS instrument, which is mounted on both the Terra and Aqua satellites, measures upwelling radiances in 36 bands with wavelengths ranging from 0.4 to 14.5µm. MODIS data, with spatial resolutions of 250, 500 m, or 1 km, have been used to construct the most detailed aerosol products, including AOD (Lee *et al.*, 2009). The most recent MODIS collection 6 (C6)

aerosol products feature enhanced Dark-Target (10 km DT) and Deep-Blue (10 km DB) AOD. The MODIS science team has carried out a few worldwide validation tests to demonstrate the cumulative impact of these adjustments and the discrepancies between the various parameters (Belle and Liu, 2016). Dark-Target (DT) was created to provide coverage over dense, dark vegetation, whereas Deep Blue (DB) was created to fill in the gaps in DT by providing coverage over bright surfaces (such as deserts) (Sayer *et al.*, 2014). In this study, the MODIS-Aqua Deep Blue AOD 550nm with a spatial resolution of 1˚ was retrieved as an average daily map from the https://giovanni.gsfc.nasa.gov, which is an online platform created by NASA for displaying and analysing geophysical parameters, with easy access to provenance.

In addition to the MODIS-Aqua AOD product, another product from another instrument is used also in this study, which is the CALIPSO (Cloud-Aerosol Lidar and Infrared Pathfinder Satellite Observations). CALIPSO's mission is an ongoing collaboration between NASA Langley Research Center (LaRC) and the Centre National D'Etudes Spatiales (CNES) to explore the global radiative effects of aerosols and clouds on climate. CALIPSO has been providing nearly continuous measurements of the vertical structure and optical properties of clouds and aerosols since its launch on April 28, 2006, to improve our understanding of their role in the Earth's climate system and the performance of a variety of models ranging from regional chemical transport to global circulation models used for climate prediction (Winker *et al.*, 2010). CALIPSO Lidar Level 1 532nm Total Attenuated Backscatter version 4.10 is the product used in this study, which describes the vertical aerosol profile and provides a clear vision about the altitude of the existing aerosols (including dust) in the troposphere and stratosphere level, more in-depth literature can be found in (Getzewich *et al.*, 2018; Kar *et al.*, 2018; Kim *et al.*, 2018). The CALIPSO 532nm Total Attenuated Backscatter images were retrieved from the official website of CALIPSO (https://www-calipso.larc.nasa.gov).

MODIS-Aqua AOD average maps were used to compare them with the average PM10 concentration maps between 0 and 100m from the HYSPLIT dust simulation results, due to the lack of PM10 ground measurements in the area of study. While CALIPSO Lidar Level 1 532nm Total Attenuated Backscatter was used to get the altitude top layer of the dust transported from the Saharan region as well as the thickness of the dust cloud over the Caribbean Sea and the South-eastern region of the United States. Also, MERRA-2 AOD data (Description of MERRA-2 AOD data can be found in section 3.4) was used for specific places where the Saharan dust particles shown to be transported to the South-eastern region of the United States to identify the intensity of the Saharan dust storm at that time in those chosen regions.

3.2.2 The identification and evaluation of the Saharan dust storm events in Budapest, Hungary between 2018 and 2022

The dust aerosol loading within the whole atmospheric column is represented by the MERRA-2 dust column mass concentration. We utilized data from the MERRA-2 Visualization tool's atmospheric composition (2D) maps (https://gmao.gsfc.nasa.gov/reanalysis/MERRA-2/data_access), as well as, hourly data (MERRA-2 M2T1NXAER V5.12.4) Obtained from NASA's Earth data website (https://disc.gsfc.nasa.gov/datasets/M2T1NXAER_5.12.4/summary).

MERRA-2 dust column mass concentration is a good measure of the intensity of the Saharan dust storm alongside with PM concentrations (Wang, Gu and Wang, 2020). In our case, to identify and evaluate Saharan dust events transported to Hungary, we used MERRA-2 dust column mass concentration data, 2D maps of dust column mass concentration, and hourly-mean PM10 mass concentrations, retrieved from Budapest Gilice tér station.

3.2.3 Case study of the Saharan dust effects on PM10 and PM2.5 concentrations in Budapest in March 2022

In 2022, Europe suffered from two severe Saharan Dust Events (SDE) during March. Large storms in March 2022 sent clouds of Saharan dust to Europe. One of them also brought long-lasting, dusty, high-altitude cirrus clouds, which caused widespread cloud cover for more than a week, from Iberia to the Arctic. It was a rare kind of storm that researchers have only recently learned to comprehend. Its characteristics include icy clouds that are infused with dust, hence the name dust-infused baroclinic storm (DIBS). A DIBS entrained and lifted an atmospheric river of Saharan dust into the troposphere in the middle of March, attaining an altitude of 10 kilometres. Dust-infused, high-altitude cirrus clouds formed as a result of the dust acting as ice nucleation particles. They continued for almost a week, covering a sizable portion of Europe. On March 15, 2022 (SDE1), the first storm developed over north-central Europe and moved south through Poland, the Czech Republic, and Austria to the eastern Mediterranean (Figure 3.5). 13 days (28th March 2022-SDE2) after the first Saharan dust storm, another wave of Saharan dust hit the south of Europe a spread to reach the Eastern European countries.

To evaluate the effect of the Saharan dust load transported to Budapest Hungary, measurements of PM10 and PM2.5 from the Hungarian Air Quality Network platform for Budapest Gilice tér air quality station.

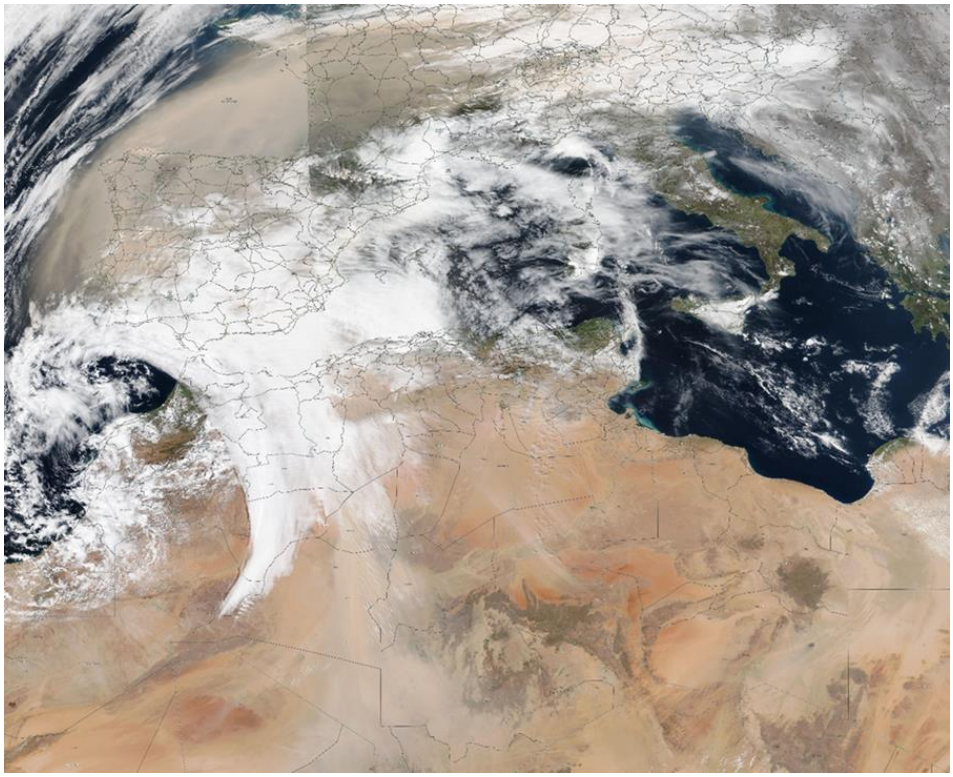

Figure 3.5. Suomi NPP / VIIRS true color image on the 15th of March 2022. Clouds appear in white and Saharan dust in pale yellow/brown

PM10 and PM2.5 concentrations are always higher in winter and fall seasons due to the alternating variability of the weather conditions and the emission source.

Figure 3.6 shows the PM10 and PM2.5 in Gilice tér air quality station during March, the first 10 days of April 2022. During March, the PM10 and PM2.5 concentrations are usually high, however, in March 2022 the PM10 concentration was below the daily EU limit value of 50 µg/m^3, alternating between 14 and 47 µg/m^3, and registering lower values in April and May.

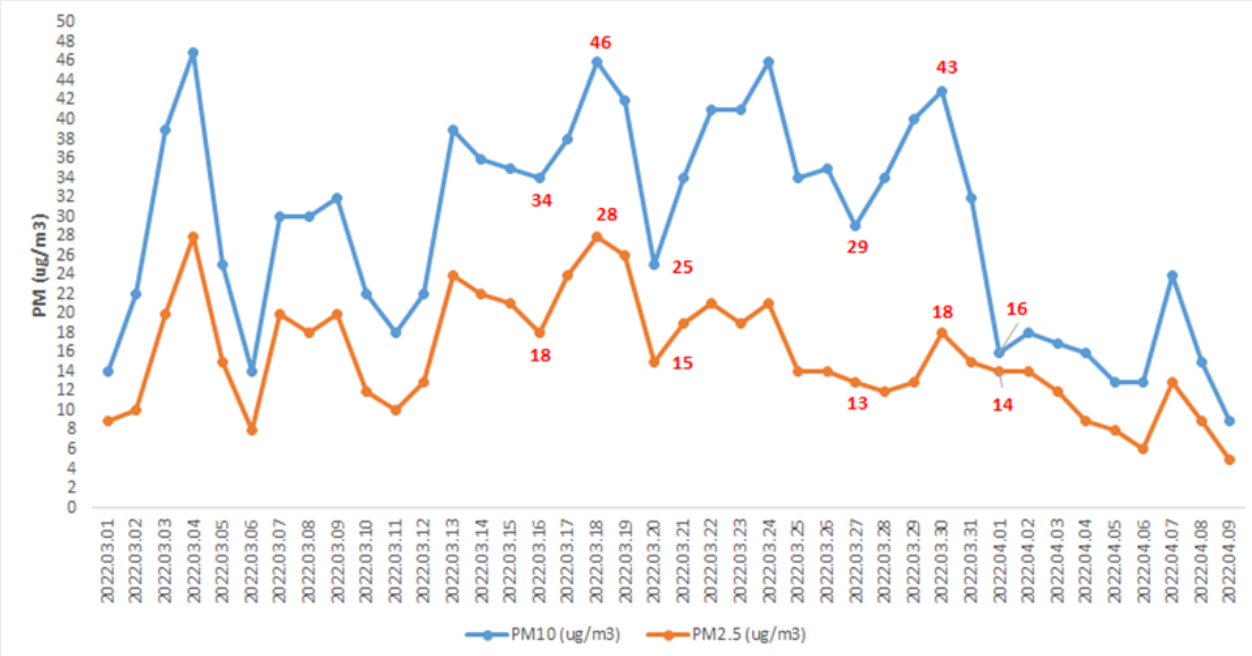

Figure 3.6. PM10, and PM2.5 concentrations (µg/m^3) in Gilice tér air quality station during March and the first 10 days of April 2022

3.3 Estimation and evaluation of PM concentrations

3.3.1 Evaluation of PM surface concentrations simulated by Version 5.12.4 of NASA's MERRA-2 Aerosol Reanalysis over Hungary in the period between 2019 and 2021

In the following sections, I describe the different methods used in Evaluation of estimated PM surface concentrations using NASA's MERRA-2 Aerosol Reanalysis over Hungary in the period between 2019 and 2021.

3.3.1.1 Description of the study

In this of the study I used two approaches. The 1st approach is estimating the PM10 and PM2.5 based on equations 1 and 2 that will be presented later and based on concentrations of components from MERRAero and compare the results with results of equations 1 and 2 with results of machine learning algorithms that will be used also to estimate PM10 and PM2.5, but based on the same concentrations of components used in equations 1 and 2 coupled with meteorological data. The second approach is estimating PM2.5 using machine learning algorithms but this time based on AOD coupled with NO_2, O_3, SO_2 and meteorological data.

The following sections will describe the data and machine learning algorithms chosen for the two approaches described before.

3.3.1.2 The MERRA-2 Aerosol Reanalysis (MERRAero)

A detailed description of the MERRA-2 Aerosol Reanalysis (MERRAero) data is provided in section 3.4.

The five PM species simulated by the MERRAero data collection every hour (SO_4, OC, BC, DS, and SS) allow for an estimation of the total PM10 concentration (Buchard *et al.*, 2016) as follows:

$$[PM_{10}] = 1.375 * [SO_4] + 1.8\,[OC] + [BC] + [DS] + [SS] \qquad (1)$$

Concentration is shown by brackets. [SO_4] is multiplied by 1.375 because it is assumed that SO_4 is completely neutralized by ammonium (NH_4) in the form of ammonium sulphate ($(NH_4)_2SO_4$). The particulate organic matter (POM) is estimated from modelled OC multiplied by a factor that

takes into account contribution from other elements associated with the organic matter. This factor has values ranging from 1.2 to 2.6 and is spatially and temporally variable (Malm *et al.*, 1994). In our simulation, a constant value of 1.8 is utilized.

Moreover, since [PM2.5], can be estimated as follows using MERRA-2 Aerosol Reanalysis data collection (Buchard *et al.*, 2016), which separates PM sizes of DS and SS:

$$[PM_{2.5}] = 1.375 * [SO_4] + 1.8\,[OC] + [BC] + [DS_{2.5}] + [SS_{2.5}] \qquad (2)$$

Equations 1 and 2 assume that SO_4, OC, and BC are all in the form of PM2.5 and do not contain nitrate particles, which can account for a sizable portion of the total [PM2.5] (Provençal *et al.*, 2017).

In our case we used AOD retrieved from MERRA-2 global atmospheric reanalysis platform for Budapest, Kecskemét and Kazincbarcika as well as in-situ measurements of PM10 and PM2.5 for the period of 2019 and 2021.

3.3.1.3 Meteorological datasets

Meteorological data were retrieved from NASA Power (Prediction of Worldwide Energy Resources) platform. The platform's list of POWER meteorological characteristics is based on the MERRA-2 assimilation model developed by NASA Goddard's Global Modeling and Assimilation Office (GMAO). Each of the parameters is either estimated using meteorological parameters acquired from NASA's MERRA-2 assimilation model, or it is directly retrieved from those values. The period from January 1, 1981, through a few months in near-real time is covered by the MERRA-2 meteorological data that is accessible through POWER. A time series of hourly (or longer time scale) values is supplied for each parameter of the POWER MERRA-2 model. The average value over the whole geographic grid is represented by each MERRA-2 parameter. The wind speed is at 10 meters, and 50 meters above the grid's average elevation and its averaged precipitation surface value. The following parameters are derived from the model at a height of 2 m above the grid box's typical elevation. The MERRA-2 parameters are computed in hourly increments and transformed to local time by the POWER project. The 24-hourly temperature measurements, not an average of those values, are used to determine the daily maximum and minimum temperatures.

In estimating PM concentrations, we used hourly temperature at 2 m (T in ˚C), wind speed at 10 m and 50 m (Ws_{10}, and Ws_{50} in m/s), Relative Humidity (RH), surface pressure (P in kPa) from NASA Power, and Planetary Boundary Layer Height (PBLH in m) from MERRA-2 global atmospheric reanalysis platform.

3.3.1.4 Machine learning algorithms

One of the finest approaches to address the complicated interaction between AOD, PM, and associated factors, such as the meteorological parameters, and typically obtain amazing predicted outcomes, is machine learning, a branch of artificial intelligence. The machine learning models were created using Python 3 and the scikit-learn library in JupyterNotebook 6.4.12.

Data preprocessing

Before applying the machine learning algorithm to the data, all data were integrated and matched by time using Microsoft Excel, and cleaned from non-values, in order to generate clean CSV file that will be loaded to the algorithm (Figure 3.7).

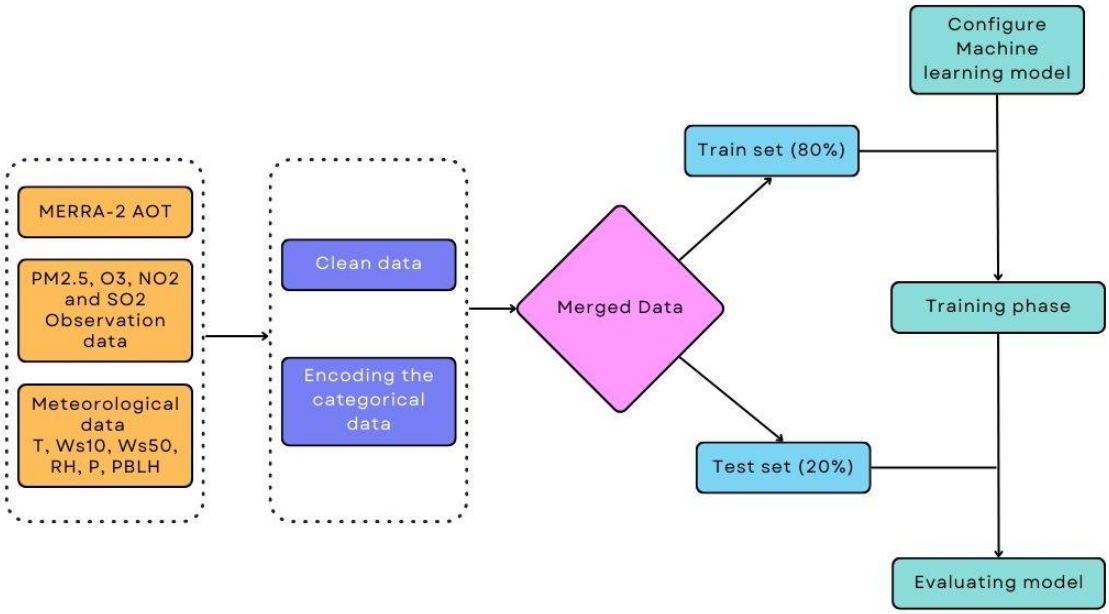

Figure 3.7. Algorithm flowchart

Multiple Linear Regression

The multiple linear regression (MLR) model is the most often used statistical tool for determining the relationship between two or more variables. MLR is an extended model of simple linear regression, where instead of using one variable to predict one outcome, multiple variables are used to predict one outcome.

MLR can be described using the following equation:

$$y = \beta_0 + \sum_{i=1}^{n} \beta_i x_i + \varepsilon \quad (3)$$

Where: y is the outcome value, x_i are the different variables, β_0 is the intercept term, β_i are regression coefficients, and ε is the error term.

Ordinary least squares regression

Ordinary least squares (OLS) models assume that the researcher is fitting a model of a relationship between one or more independent variable and a continuous or at least increment outcome variable that reduces the sum of square errors, in which an error is the difference between the real and predicted value of the outcome. Linear regression (with a single or many predictor variables) is the most frequent analytical approach that employs OLS models (Michalos, 2014).

OLS regression is increasingly employed in many scientific disciplines, including physics, economics, and psychology, and a variety of textbooks have been created to teach this approach and its applications in many fields of inquiry (Cohen, 2013; Kleinbaum *et al.*, 2013; Montgomery, Peck and Vining, 2020).

Random Forest regression

One of the most well-liked ensembles learning strategies based on decision tree predictors is Random Forest (RF), which is a straightforward, effective, and understandable strategy. The trees are bagged in the first stage, and then the tree is divided in the second step using the random subspace technique or the random split selection, applied at each node of the algorithm, and

utilizing just a subset of the characteristics to split the node. The benefits of RF included completing regression and classification tasks as well as generating accurate predictions and outcomes that can be simply explained (Breiman, 2001).

Extra Tree regression

The Extra-Trees approach (XT) employs the traditional top-down construction method to construct an ensemble of unpruned decision or regression trees. It separates nodes by selecting cut-points completely at random, which sets it apart from previous tree-based ensemble approaches. It also grows the trees using the entire learning sample rather than a bootstrap replica. In the worst situation, XT creates completely random trees, whose architectures are independent of the learning sample's output values. By selecting the right parameter, the power of the randomization may be adjusted to the particulars of the situation. The algorithm's biggest advantage, aside from accuracy, is computational speed (Geurts, Ernst and Wehenkel, 2006).

In the first approach, we estimated the PM10 and PM2.5 using MLR, OLS, RF and XT machine learning algorithms, based on BC, OC, DS, SS, SO_4, AOD, and PBLH from MERRAero data, and T, RH, Ws_{10}, Ws_{50} and P from NASA Power platform. While in the second approach, we estimated the PM2.5 using MLR, OLS, RF and XT machine learning algorithms, based on AOD, and PBLH from MERRAero data, and T, RH, Ws_{10}, Ws_{50} and P from NASA Power platform, and measurements from Hungarian air quality network of NO_2, SO_2 and O_3. In addition, both approaches were done for Budapest, Kecskemét and Kazincbarcika for the period of 2019 and 2021.

3.3.2 Calibration of CAMS PM2.5 data over Hungary using machine learning

The purpose of this study is to calibrate CAMS PM2.5 data using the LightGBM algorithm and evaluate its impact on improving the accuracy and correlations with in-situ measurements in Hungary. The study aimed to address the limitations of raw CAMS data and provide more reliable information for air quality assessments.

3.3.2.1 CAMS global reanalysis (EAC4)

The Copernicus Atmosphere Monitoring Service (CAMS) reanalysis is the most recent global reanalysis dataset of atmospheric composition produced by the European Centre for Medium-Range Weather Forecasts (ECMWF), and it consists of three-dimensional time-consistent atmospheric composition fields, which include aerosols and chemical species (Inness *et al.*, 2019). The CAMS reanalysis expands on the knowledge gathered during the previous Monitoring Atmospheric Composition and Climate (MACC) reanalysis and interim CAMS reanalysis. Total column CO, tropospheric column NO_2, aerosol optical depth (AOD), and total column, partial column, and profile ozone retrievals from satellites were used in the CAMS reanalysis with the ECMWF's Integrated Forecasting System. The CAMS forecasts air pollution levels throughout the world over the following few days. The CAMS, in particular, generates a forecast of global atmospheric composition with time horizons as long as the next 120 hours, consisting of 56 reactive trace gases in the troposphere, stratospheric ozone, and five different types of aerosols (i.e., desert dust, sea salt, organic matter, black carbon, and sulphate) (Wagner *et al.*, 2021).

EAC4 (ECMWF Atmospheric Composition Reanalysis 4) is the fourth generation of ECMWF global atmospheric reanalysis. Reanalysis integrates model data with observations from throughout the world to create a globally complete and consistent dataset using an atmosphere model based on physical and chemical rules. This data assimilation principle is based on the method employed by numerical weather prediction centers and air quality forecasting centers (Peuch *et al.*, 2018). The assimilation system can estimate biases between observations and separate high-quality data from low-quality data. Estimates can be made using the atmosphere

model in areas with limited data coverage or for atmospheric contaminants for which no direct observations are available. Reanalysis is a very convenient and popular dataset to work with since it provides estimates for each grid point around the world for each regular output time over a long period of time, always in the same format. The observing system has evolved significantly over time, and while the assimilation system can fill data gaps, the initially more sparser networks result in less accurate estimations. As a result, EAC4 is only available since 2003 (Copernicus, 2020). CAMS gives global estimates every 3 h, with a horizontal resolution of 0.75°x 0.75° and a vertical structure of 60 hybrid model levels, with a top-level at 0.1 hPa.

In the current study we used a single-level PM2.5 data downloaded from CAMS website (https://ads.atmosphere.copernicus.eu) for the years 2019 and 2020.

3.3.2.2 ERA5 Meteorological datasets

ECMWF prepared the ERA5 reanalysis as part of the Copernicus Climate Change Service (C3S), which will contain a full record of the global atmosphere, land surface, and ocean waves from 1950 onwards. This new reanalysis will take the place of the ERA-Interim reanalysis, which began in 2006 (Hersbach *et al.*, 2020). ERA5 produces hourly estimates for a wide range of atmospheric, oceanographic, and land-surface variables. An underlying 10-member ensemble samples an uncertainty estimate at three-hourly intervals. For your convenience, the ensemble mean and spread have been pre-calculated. Such uncertainty estimations are intimately tied to the available observing system's information content, which has developed significantly over time. They also show flow-dependent sensitivity zones. Monthly-mean averages have also been pre-calculated to help with many climatic applications, while monthly means for the ensemble mean and spread are not available. For the reanalysis, data was regridded to a standard lat-lon grid of 0.25 degrees and 0.5 degrees for the uncertainty estimate (0.5 and 1 degree for ocean waves, respectively). There are four major subsets: hourly and monthly products on pressure levels (upper air fields) as well as single levels (atmospheric, ocean-wave, and land surface values). ERA5 hourly data on single levels starts from 1940 to the present (Hersbach *et al.*, 2020).

Data was downloaded from Copernicus climate data platform website (https://cds.climate.copernicus.eu). In this study we used the temperature of air at 2 m above the surface (T in ˚C), relative humidity (RH), Planetary boundary layer height (PBLH in m), 10 m u and v components of wind (u10 and v10 in m/s), surface pressure (P in Pa) and total could cover (tcc).

3.3.2.3 LightGBM algorithms

LightGBM is a highly effective and scalable gradient boosting decision tree technique that benefits from its histogram-based approach, leaf-wise tree development strategy, and proprietary feature bundling (Ke *et al.*, 2017). LightGBM algorithms are a type of gradient boosting framework that have received a lot of attention due to their remarkable performance and efficiency when dealing with large-scale datasets (Sheridan, Liaw and Tudor, 2021). LightGBM algorithms help to advance cutting-edge technology by boosting our understanding of complicated data patterns and, eventually, decision-making processes across numerous industries. Overall, LightGBM algorithms offer extraordinary societal benefit by expanding the field of machine learning and enabling more accurate and efficient data processing (Xia *et al.*, 2021).

LightGBM can process massive amounts of high-dimensional big data with greater efficiency and performance than traditional machine learning approaches. In our study, LightGBM is an appropriate choice. The mathematical equations for PM2.5 calibration schemes are as follows:

$$PM_{2.5,C} = f_{model}(\ CAMSPM2.5, T, RH, blh, u10, v10, P, tcc, hour, day, month) \qquad (4)$$

The data preprocessing and data matching phase involved preparing and aligning the CAMS and ERA5 datasets for further analysis. The resolution of the CAMS dataset is 0.75x0.75, while the resolution of the ERA5 dataset is 0.25x0.25. A geographic matching procedure was used to match the air quality stations with the relevant grid points in each dataset. The purpose was to find the CAMS and ERA5 grid point that was nearest to each air quality station. The geographic coordinates of the air quality stations were matched with the grid points in both the CAMS and ERA5 datasets during the data matching process. The closest grid point to each station was obtained by computing the distances between the station coordinates and the grid point coordinates.

After the data preprocessing and matching phase, the datasets were further divided into training and test sets (80 x 20% split) with 5-fold cross validation. The Pearson correlation R is calculated between the raw CAMS and in-situ PM2.5 before the training of the model for the test data, and after of the training of the model between calibrated and in-situ PM2.5.

3.4 Data and statistics

3.4.1 The MERRA-2 Aerosol Reanalysis (MERRAero)

The Goddard Earth Observing System Model, Version 5 (GEOS-5) is the foundation of the MERRA-2 assimilation system (Molod et al., 2015). MEERA-2 incorporates spaceborne aerosol products from Moderate Resolution Imaging Spectroradiometer (MODIS), Multi-angle Imaging Spectro Radiometer (MISR), and the ground-based remote sensing network AErosol RObotic NETwork (AERONET) as data for its aerosol dataset. The optical characteristics, emissions, deposition, and aerosol mixing ratios of the five different types of aerosols are all included in the MERRA-2 aerosol dataset vertically (Buchard *et al.*, 2017; Randles *et al.*, 2017). The data from MERRA-2 comprise 21 different types of products, such as atmospheric aerosols, radiation, temperature, water vapor, precipitation, etc. The data span the years 1980 to the present, and are saved in a standard grid of 0.5° × 0.625° (Randles *et al.*, 2017).

The GOCART (the Goddard Chemistry Aerosol Radiation and Transport model) chemistry module, which simulates five different forms of aerosols, is integrated with the MEERA-2 model (sulfate (SO_4), organic carbon (OC), black carbon (BC), sand dust (DS), and sea salt (SS)). These aerosols are considered as external mixes that do not interact with one another. While the surface wind speed affects the emissions of dust and sea salt, other aerosol types are predicted from potential natural and anthropogenic sources. Convective large-scale wet removal, dry deposition, sedimentation, and chemical processes to generate sulphate aerosol from Sulphur dioxide (SO_2) oxidation are all included within the GOCART model (Randles *et al.*, 2017).

The parameterizations of natural and anthropogenic emissions in MERRAero have got numerous significant modifications from the previous edition of the GEOS-4 modelling system (Colarco et al., 2010). The Edgar-4.1 inventory was used to calculate SO2 emissions from anthropogenic sources, and the injection scheme was changed to account for changes in the injection profiles of emission sources from the energy and non-energy sectors (Buchard *et al.*, 2014). The emissions from biomass burning are from the NASA Quick Fire Emission Dataset (QFED) Version 2.1. QFED is a worldwide fire radiative power-based inventory of daily aerosol precursor and trace gas emissions (Koster, Darmenov and da Silva, 2015). According to the study of Jaeglé *et al.* (2011) a novel independently obtained sea surface temperature (SST) adjustment term was used to modify the intensity of sea-salt emissions. Dust emission is predicated on the correlation of reported dust source sites with large-scale topographic depressions, as proposed by Ginoux *et al.* (2001).

MERRA-2 coupled AOD at 550 nm, from a variety of ground- and space-based remote sensing platforms, including (i) bias-corrected AOD from Moderate Resolution Imaging

Spectroradiometer (MODIS) Terra and Aqua, (ii) the Advanced Very High Resolution Radiometer (AVHRR) instruments, (iii) AOD retrievals from the Multiangle Imaging SpectroRadiometer (MISR) over bright surfaces, and (iv) ground-based Aerosol Robotic Network (AERONET) direct measurements of AOD (Level 2) (Randles *et al.*, 2017).

3.4.2 Air quality stations

The Hungarian Air Quality Monitoring Network provides real-time and historical air quality monitoring data throughout Hungary. The network is divided into two main parts: automatic monitoring stations that continuously measure a wide range of air pollutants in the ambient air, and a manual system with sample points and subsequent laboratory examination. The existing network in Hungary comprises 37 fully automatic monitoring stations. The National Air Quality Reference Centre and Laboratory's primary responsibilities are as follows: Oversight of the operation of the Hungarian Air Quality Monitoring Network (HAQM) in accordance with Ministry of Agriculture standards, coordination and regulation of HAQM methods and procedures in accordance with EU regulations, maintain measurement traceability by running an approved Calibration Laboratory, and participation in national and worldwide standards development. A CO analyser, PM10 / PM2.5 monitors, a calibration tower, and a mass flow meter calibration system were added to the calibration laboratory instrument fleet (Weidinger *et al.*, 2010).

Among the monitoring sites in Budapest, the Gilice tér urban background station (located in the SE part of the city) was chosen for our analysis because it is a standard meteorological and air quality monitoring station that provides PM10 and PM2.5 concentrations and detailed meteorological observations with good data coverage.

Kecskemét is located 86 kilometers from both the capital Budapest and the country's third-largest city, Szeged, and is almost equal distance from the country's two major rivers, the Danube and the Tisza. Kecskemét is the city most vulnerable to climate change, with a slew of environmental issues in the Danube-Tisza Interfluve. The most significant changes include the degradation of air quality, the influence of urban heat islands, and water management (Hoyk, Kanalas and Farkas, 2020). The air quality station in Kecskemét is an urban background station.

Kazincbarcika is a town in the county of Borsod-Abaj-Zemplén in Northern Hungary. It is located in the valley of the Sajó River, 20 km away from Miskolc, the county capital. The air quality station in Kazincbarcika is an international urban background station. Table 3 presents the list of air quality stations used throughout the different studies as well as their geographical coordinates.

All PM10 and PM2.5 data were retrieved from the Hungarian Air Quality Network platform (Országos Légszennyezettségi Mérőhálózat (OLM), https://legszennyezettseg.met.hu), which is a platform that provides actual and historical air quality monitoring data throughout Hungary.

Table *3*: List of air quality stations with latitudes and longitudes

Station	**latitude**	**longitude**
Ajka	47.10	17.55
Budapest Gilice	47.43	19.18
Kazincbarcika	48.24	20.61
Kecskemet	46.90	19.68
Miskolc_Alfoldi	48.09	20.81
Nyiregyhaza	47.96	21.71
Pecs Nevelesi Kozpont	46.04	18.22
Szazhalombatta_Buzavirag_ter	47.31	18.92
Szeged_Rozsa	46.27	20.15
Szolnok	47.18	20.2
Veszprem	47.09	17.9

3.4.3 Performance statistics

The performance of the air quality forecast models (in sections 3.3.1 and 3.3.2) using the testing dataset was assessed using model performance metrics, such as R^2 computed by Equation (5), RMSE calculated by Equation (6), MAE calculated by Equation (7), and Pearson correlation R calculated by Equation (8)

$$R^2 = \frac{\left[\int_{i=1}^{n}(p_i - \bar{p}) - (o_i - \bar{o})\right]^2}{\left[\int_{i=1}^{n}(p_i - \bar{p})^2\right]\left[\int_{i=1}^{n}(o_i - \bar{o})^2\right]}, \quad (5)$$

$$RMSE = \sqrt{\sum_{i=1}^{n}\frac{(p_i - o_i)^2}{n}}, \quad (6)$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|p_i - o_i|, \quad (7)$$

$$R = \frac{\sum(p_i - \bar{p}) - (o_i - \bar{o})}{\sqrt{\sum(p_i - \bar{p})^2 \sum(o_i - \bar{o})^2}} \quad (8)$$

Where:

p_i the predicted value of the sample, and $\bar{p}$ is the predicted average.

o_i the observation value, and $\bar{o}$ is the observation average.

n the number of the samples.

4 RESULTS

In this chapter, I present the results of all the 3 main chapters presented in Material and Method section.

4.1 PM dispersion experiments

4.1.1 Small scale experiments of PM10 dispersion around obstacles

The results of the experiments showed some interesting aspects for the understanding of the PM10 dispersion around simple obstacle (Wall).

4.1.1.1 Sensor A

The sensor A is the sensor behind obstacle. Figure 4.1 shows the average concentration of PM10 during each experiment in function of Obstacle heights (OH) and distance from the source (OD). The average PM10 concentration increase with increasing of the obstacle distance from the source at higher wind speed while in low wind speed it is almost stable. At wind speed of 2.9 m/s the average PM10 concentration was the same for obstacle height 240 and 360 mm while it was at its peak when obstacle height was 120 mm. while, for wind speed of 1m/s the peak average PM10 concentration was at obstacle height of 360 mm and almost the same in the other two heights.

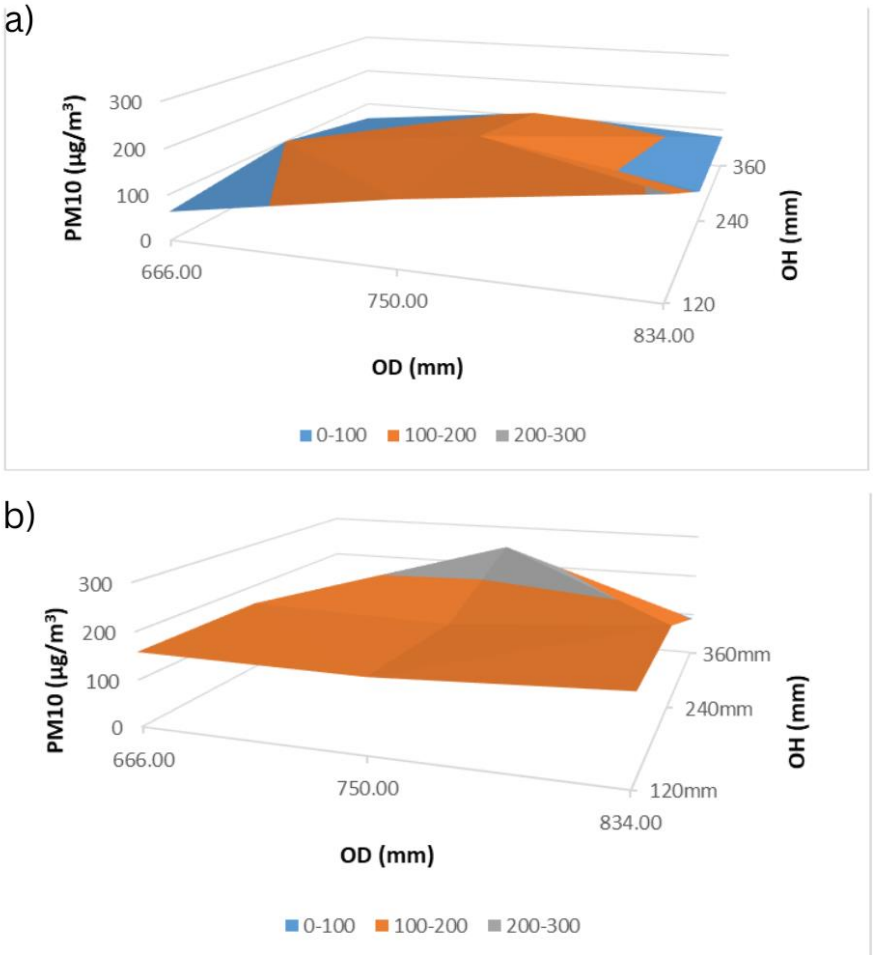

Figure 4.1. graphs of Average PM10 concentration registered by Sensor A in function of Obstacle heights and distance from the source in case of a) wind speed 2.9 m/s and b) wind speed 1 m/s

4.1.1.2 Sensor B

For the sensor B (Figure 4.2), which is the sensor placed before the wall, the PM10 average concentration was higher in case of wall height of 240 and 360 mm, and wall distance of 750 mm

at wind speed of 1 m/s. While it reaches the maximum when obstacle distance from the source is 834mm, obstacle height is 120 mm and wind speed of 2.9 m/s.

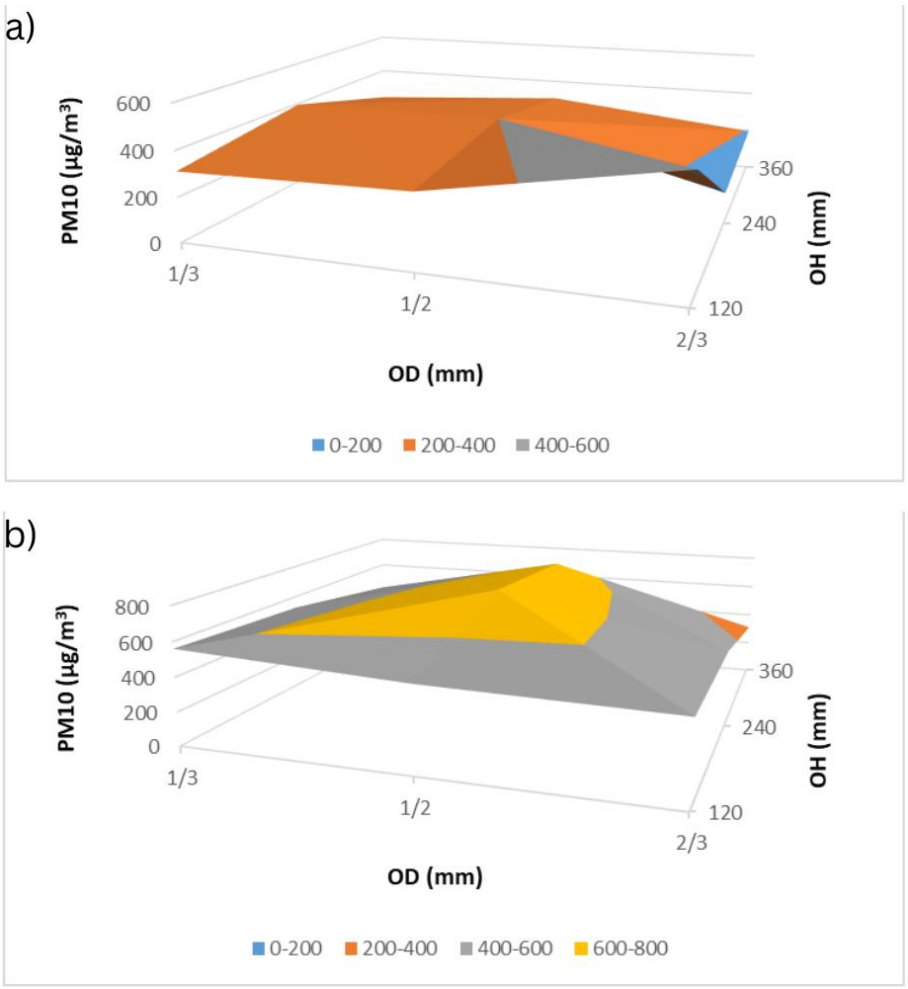

Figure 4.2. Graphs of Average PM10 concentration registered by Sensor B in function of Obstacle heights and distance from the source in case of a) wind speed 2.9 m/s and b) wind speed 1 m/s

4.1.1.3 Sensor C

The sensor C placed near the source registered almost same average concentration of PM10 at wind speed of 1m/s with decrease in concentration in case of obstacle height 360 mm and distance from source 834 mm (Figure 4.3). In the other hand it was changing at wind speed of 2.9 m/s. The peak average PM10 concentration was as the same as sensor B, when obstacle distance from the source is 834 mm and obstacle height is 120 mm.

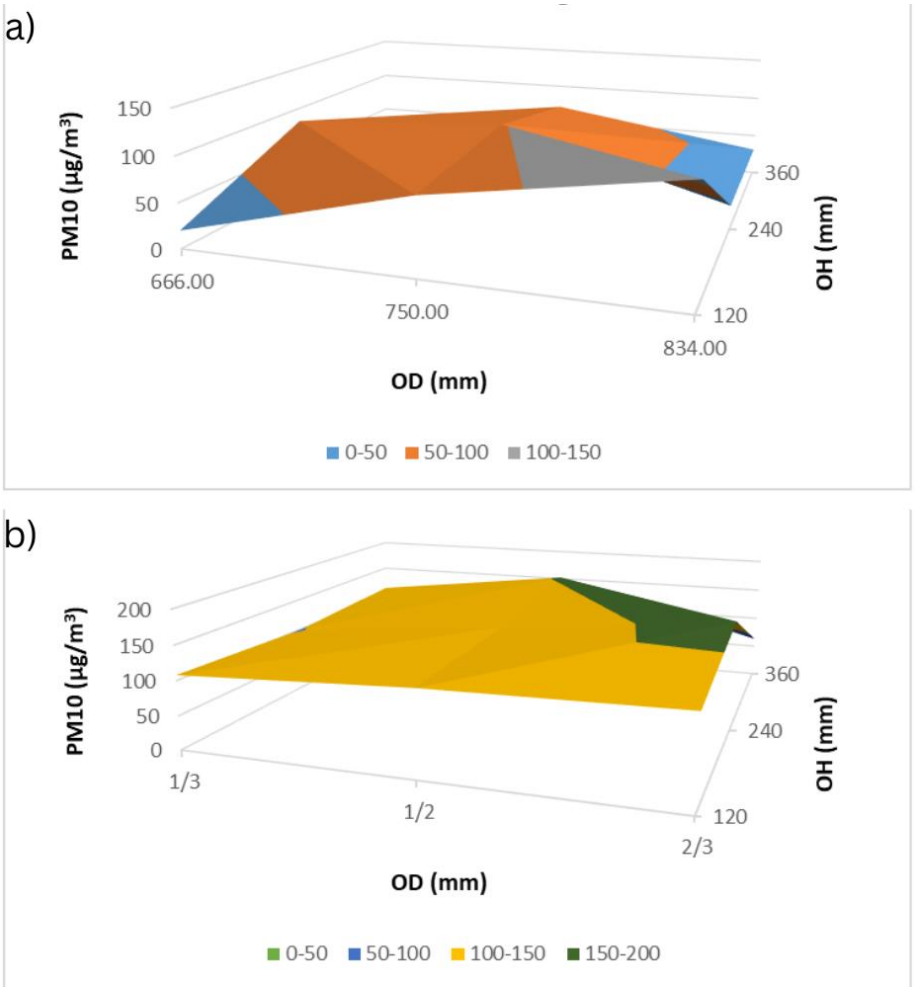

Figure 4.3. Graphs of Average PM10 concentration registered by Sensor C in function of Obstacle heights and distance from the source in case of a) wind speed 2.9 m/s and b) wind speed 1 m/s

The result of a multiple regression analysis aims to predict $PM10_A$ concentration using four independent variables: "OH", "OD", "$PM10_C$," and "Ws". The resulting the equation of the regression model (Eq.9) can be written as follows:

$$PM10_A = 143.07 - 71.86 * OH - 171.42 * OD + 1.23 * PM10_C + 12.34 * Ws \quad (9)$$

The correlation coefficient (R) was 0.89, suggesting that the dependent variable and the set of independent factors had a moderately significant positive connection. Furthermore, the coefficient of determination (R^2) was 0.79, indicating that the independent variables in the model explain roughly 79% of the variation in the dependent variable.

The results of this research concluded that there is a positive significant effect of Obstacle heights, distance of the obstacle from the source, and the wind speed. The PM10 average concentration decrease significantly in the sensor A (behind the wall) when the obstacle height increases and also when the obstacle distance from the source increase also in case of the two-wind speed (1 m/s and 2.9 m/s) with higher concentrations registered in case of wind speed is 1 m/s. while, changes in the PM10 average concentration was also seen in case of Sensor B (in the middle) and sensor C (near the source) especially in case of high wind speed (2.9 m/s) and that is due to the turbulence created before and after the walls when the wind hits it, in addition to the reflexing of PM plumes by the obstacle. Also, maximum PM10 concentration sensor A (after wall) and sensor B (before wall) at obstacle distance 834 mm, and obstacle height 120 mm, while at low wind speed (1 m/s) the PM10 concentrations does not change with effects of obstacle height and distance from source.

In contrast, at higher wind speed (2.9 m/s), the obstacle height and distance affect the PM10 concentration before and after the obstacles in the same way, meaning that the concentrations tend to decrease with obstacle height increases, and as close as the obstacle to the sensor the concentration increases with low obstacle height, due to the turbulence created near the obstacle which trap the PM10 particles near the obstacle. Thus, the experiments results prove the same effects of simple obstacle presence as larger scale study where complex urban landscape and structure are involved. The experiments proved that also in small scale experiments the transportation of the PM particles are the same as in real scale transportation of PM.

Generally, the PM10 average concentration tends to decrease when obstacle heights increase but also combined with position of the obstacle far from the source. In our case, the experiment is a simplification of the dispersion of PM concentration (PM10 specifically) in an austere environment. It represents the basis for understanding the PM pollutant source interaction with the barrier and how it affects PM10 concentrations. The results may change in a complex urban setting, where many parameters can intervene to change the dispersion of air pollutants. Our case study's results are valid but subject to investigation in other experimental settings.

Moreover, using Incense sticks as source of PM pollution showed that while the stick is burning it continues to spike the PM10 concentration, as before the experiments the background concentration of PM10 was 7±3 $\mu g/m^3$ and during the experiments it can reach 700 $\mu g/m^3$, which manifest the short-term effect of burning the incense stick and its risk of affecting the indoor air quality if used in excess. Finally, the experiment is representation of trying to find simple obstacle placement that can reduce significantly PM plume coming from source that could be industrial or traffic source. The results show that the higher the obstacle is better but also closer is better also, but in real situation simple obstacle can be put in the way of PM plumes and as closer as possible to the area that is subject to be defended from high PM concentrations. And one of the best options is to combine simple obstacle (solid barrier) with vegetated/tree barrier as the last was proven to improve air exchange, and The tree planting and trunk height have a considerable impact on the air flow and pollution dispersion (Buccolieri *et al.*, 2022).

4.1.2 Effect of small hills on PM10 and PM2.5 concentrations in short range

The average concentrations registered by sensor 3 (S3) of PM10 and PM2.5 are higher in the case of the 1m height and 0.8m height compared to the concentrations recorded during flat case Figure 4.4.

At low wind speeds (0 and 0.7 m/s), the average concentrations of PM10 and PM2.5 registered by S3 are almost the same in all the 3 cases. At wind speeds of 2.4, 3.7, and 5.1 m/s, the average concentration of PM10 and PM2.5 are higher in the case of the two different heights compared to flat areas. The peak concentration of PM10 and PM2.5 in case of 1m height registered when the wind speed was 3.7 m/s, while in case of 0.8 m height was at a wind speed of 5.1 m/s, while in a flat area, average concentrations registered were almost the same when wind speed was higher than 2.4 m/s. In addition, the same in the case of 0.8 height, but the average concentration was 2.5 to 3 times higher than in the flat case with a slight decrease at high speed (6.1 m/s). While in the case of 1m height, the average PM concentration was 2 to 3 times with wind speeds of 3.7 and 5.1m/s, and almost the same at wind speeds of 2.4 and 6.1 m/s.

The difference in the ground surface elevation between case 2 and case 3 is just 0.2 m, but the effect on the dispersion of the PM plumes can be seen from the average PM concentrations. In the case of a flat ground surface, the spread of PM pollutants is parallel to the wind direction. In contrast, high ground (in our case, in the form of a hill) at different elevations changes the dispersion pathway of the PM particles. The different slopes of the hills create other flows of the

PM dispersion; in case two, the approximate same PM concentrations registered in different ranges of wind speed means the PM particles are trapped in the same way regardless of the wind speed.

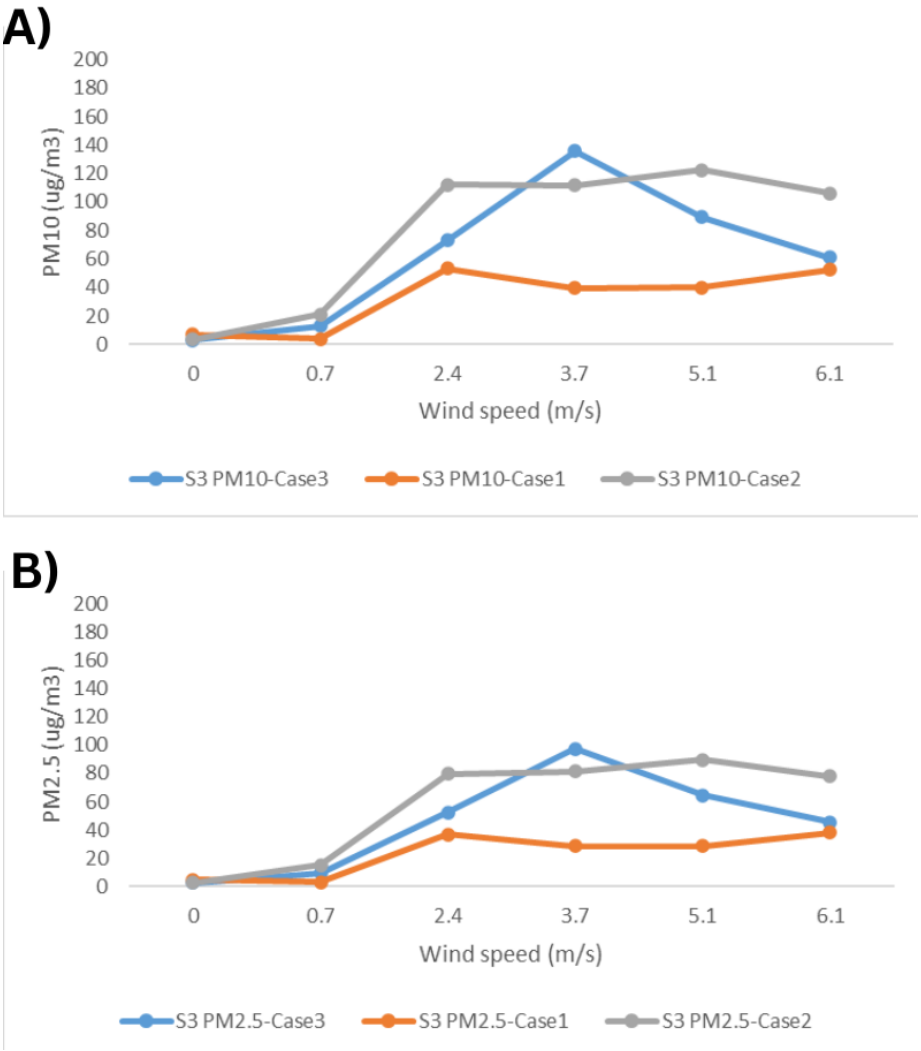

Figure 4.4. Average S3 concentrations in all cases of A) PM10 and B) PM2.5

Comparing the three PM10 concentrations registered by the 3 sensor (Figure 4.5), it's seen that S1 registered higher concentrations in case 2 and 3 than in case 1, especially at wind speeds less than 3 m/s, and that is due to reflective effect of the hill and also low wind speed. While for S2 the PM particles are trapped before the hill which promote higher PM concentrations.

In this study also, multiple linear regression method was used to estimate PM10 concentration at the top of the hill ($PM10_{S3}$ in µg/m3) based on the PM10 concentration near source (concentration registered by Sensor 1, $PM10_{S1}$), PM10 concentration at the bottom of the hill (concentration registered by Sensor 2, $PM10_{S2}$), the wind speed (Ws in m/s), and the height of the hill (H in m).

The result of the multiple linear regression is the following equation:

$$PM10_{S3} = 5.92 - 0.173 * PM10_{S1} + 0.580 * PM10_{S2} + 4.29 * Ws - 11.29 * H \quad (10)$$

The correlation coefficient (R) was 0.9, indicating a relatively strong positive correlation between the dependent variable and the combination of independent variables. In addition, the coefficient of determination (R^2) was 0.82, which means that approximately 82% of the variance in the dependent variable is explained by the independent variables in the model.

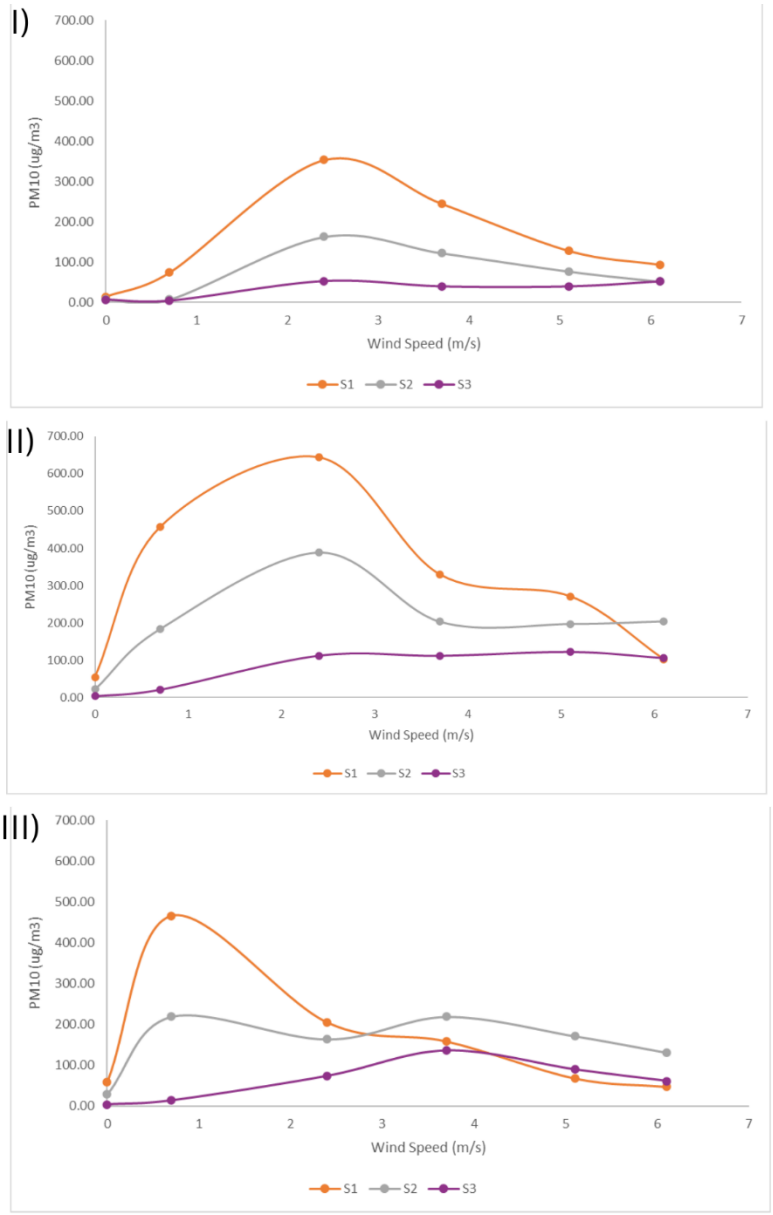

Figure 4.5. PM10 concentrations of the 3 sensors in I) case 1, II) case 2, and III) case 3

Analysing the correlations between PM10 concentrations of sensors S1, S2 and S3 (Figure 4.6), it shows that for S1 correlation was low and positive in case 1 (0.2), but it changes to negative in cases 2 and 3 (-0.18 and -0.5 respectively), which show the effects of the height of the hills. For S2, the correlation between PM10 concentration and wind speed decrease as the height of the hill increase, while for S3, a strong correlation is observed in case 1 and 2 (0.8 and 0.84, respectively), and it decreases in case 3. Thus, the decrease in the correlation due to the higher elevation of the hill could be because of the changes in the wind flow created by different elevations of the hill. The results underscore the significant influence of hill elevation on the correlation between PM10 and wind speed at various sensor locations, emphasizing the role of local topography in shaping air pollution patterns during the experiments.

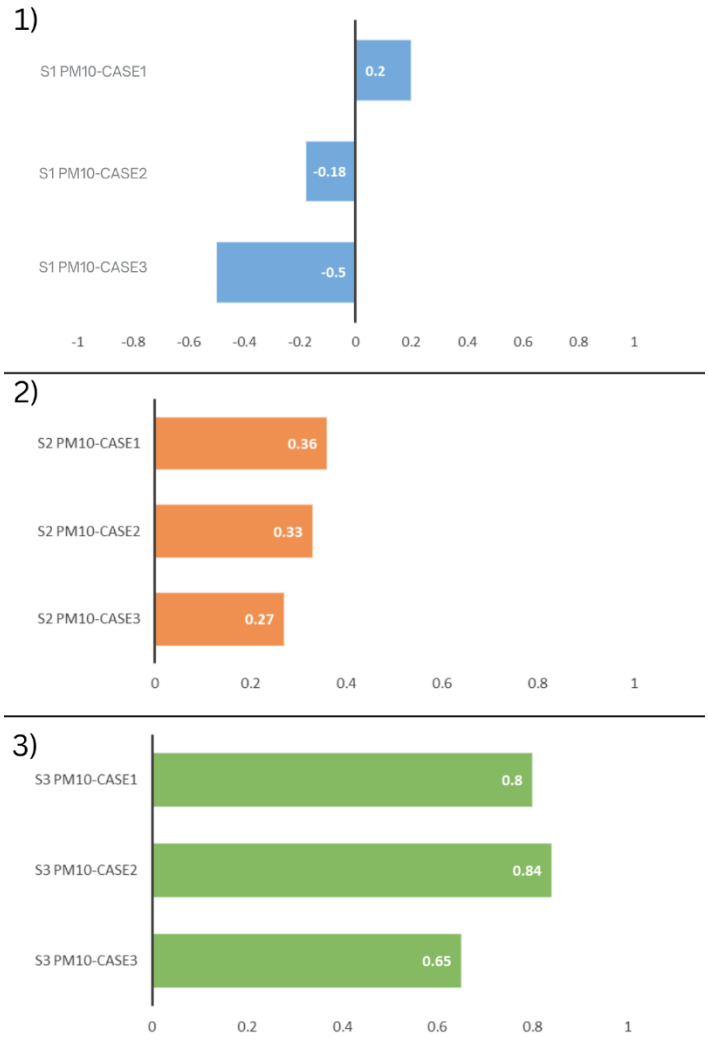

Figure 4.6. Correlations between PM10 concentrations and wind speed in all cases for 1) Sensor 1, 2) Sensor 2, and 3) Sensor 3

4.2 Saharan Dust storm transport

4.2.1 Dust Storm simulation over the Sahara Desert (Moroccan and Mauritanian regions) using HYSPLIT

In this chapter I describe the results Saharan Dust simulation study using HYSPLIT model.

4.2.1.1 Dust Simulation and cluster analysis results

Hysplit PM10 emission modelling results (Figures 4.7 and 4.8) show that the PM10 emission on the 14th of June 2020 started from the region of Tinduf, Algeria (Close to the Moroccan borders), Adrar, Tiris Zemmour, and Tagant in Mauritania. While the dust storm was continuous for 4 days and the dust was transported to the North Atlantic Ocean, the average PM10 concentration between 0 and 100m was between 100 µg/m^3 and 10000 µg/m^3 in some critical regions like Tinduf, Algeria on the 14th and 17th of June 2020, Adrar, Mauritania on the 15th, Bir Anzarane, Morocco on the 16th, Tiris Zemmour, Mauritania on the 17th, Goundam Cercle, Mali on the 18th of June 2020. Comparing the average PM10 concentration maps between 0 and 100m from the HYSPLIT modelling results and the MODIS Aqua Deep Blue AOD maps (Figure 4.9), it can be seen that in most of the regions where the PM10 concentration are high, the AOD index is also at a high level, which indicates a positive correlation between the PM10 concentration and the AOD index of MODIS Aqua. Regions like Tiris Zemmour in Mauritania, Western Sahara of Morocco, Western

and Southern regions of Algeria, are also characterized as source regions that influence the level of the PM10 concentration over the western Mediterranean Basin (Salvador et al. 2014; Russo et al. 2020). Moreover, all the areas that had a high concentration of PM10 in the HYSPLIT dust simulation results and high AOD values (between 0.7 and 1) according to time-averaged maps of MODIS-Aqua are considered as primary dust natural source regions (Ginoux et al. 2012), and they are active throughout the year, although their peak activity is between April and September (Prospero et al. 2002).

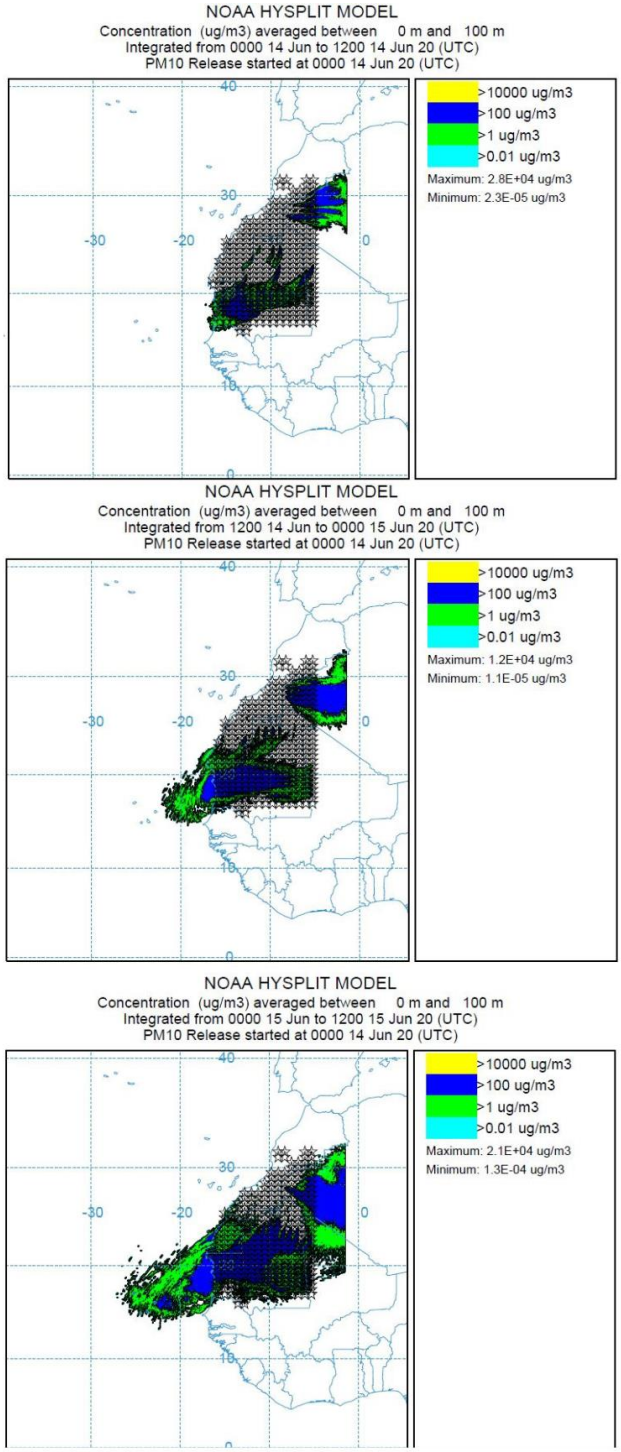

Figure 4.7. Modelling results for the concentration of PM10 averaged across the 0-100m altitude range in June a) 14th from 00 UTC to 12 UTC, b) 14th from 12 UTC to 15th 00 UTC c) 15th from 00 UTC to 12 UTC

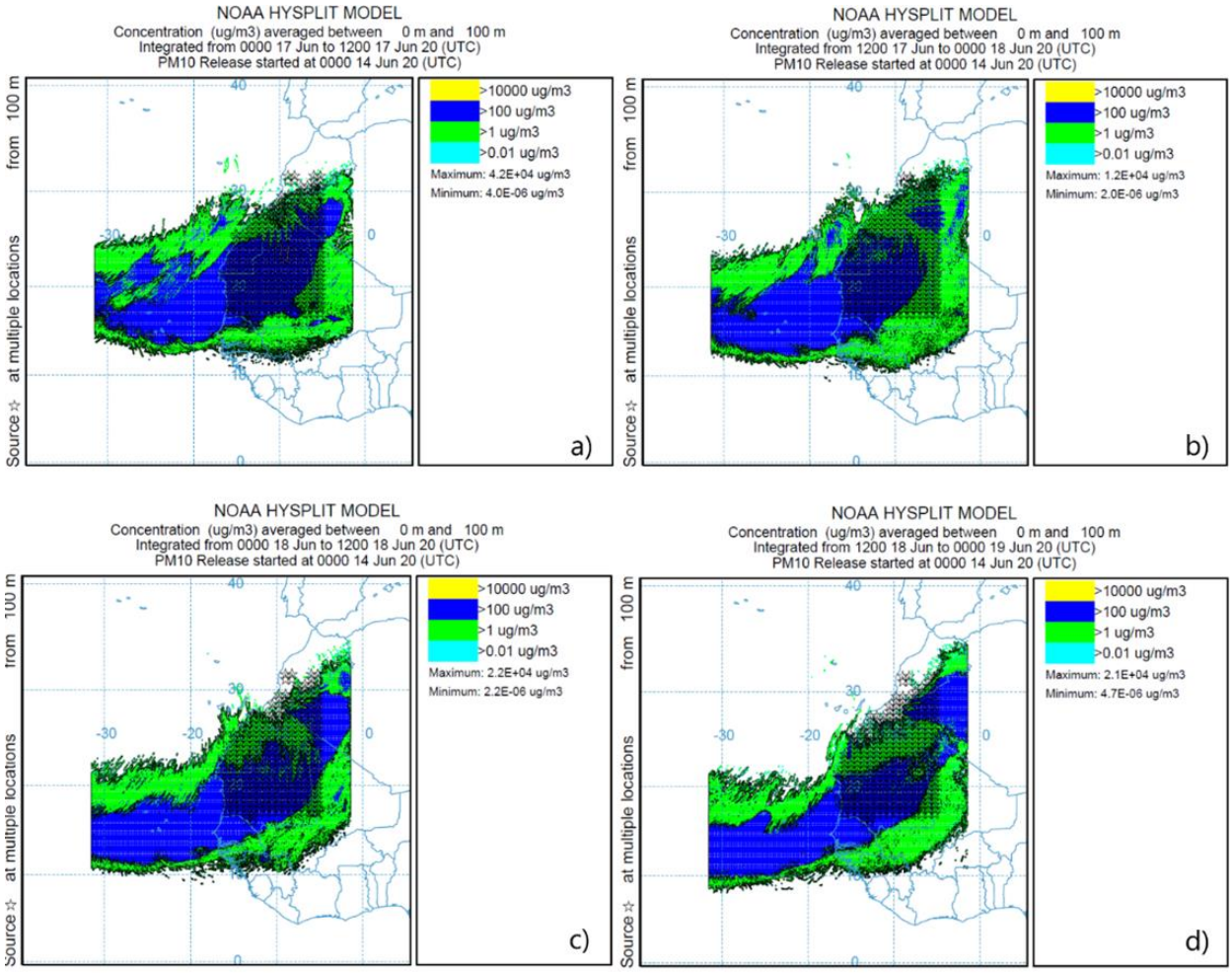

Figure 4.8. Modeling results for the concentration of PM10 averaged across the 0-100m altitude range in June a) 17th from 00 UTC to 12 UTC, b) 17th from 12 UTC to 18th 00 UTC c) 18th from 00 UTC to 12 UTC d) 18th from 12 UTC to 19th 00

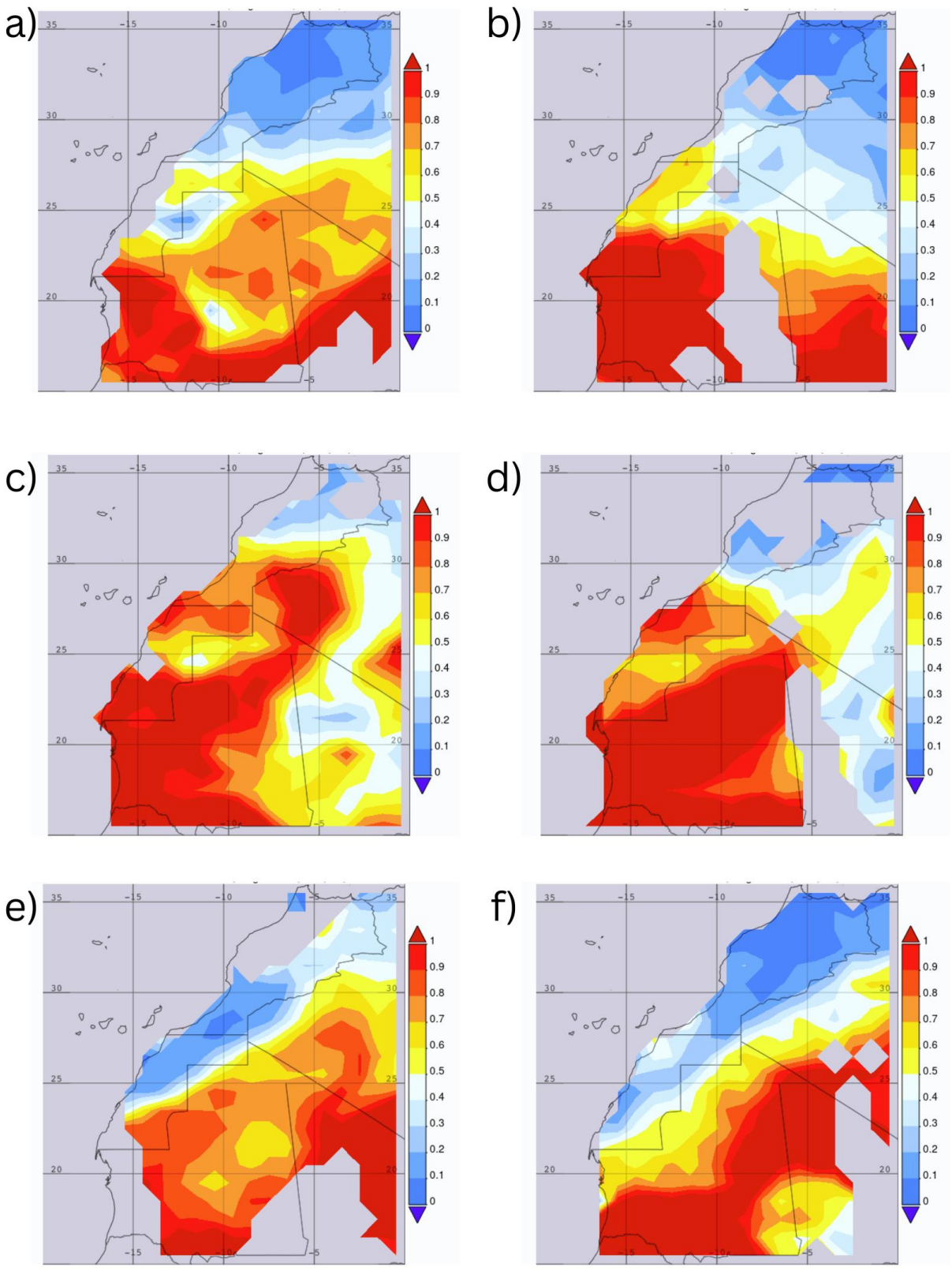

Figure 4.9. MODIS Aqua time Averaged Map of Aerosol Optical Depth 550 nm (Deep Blue, Land-only) daily 1 deg in the region of western Sahara on the a) 14, June 2020 b) 15, June 2020 c) 16, June 2020 d) 17, June 2020 e) 18, June 2020 f) 19, June 2020

The analysis of the trajectories of the PM10 particles emitted from numerous locations in the western Sahara during the June dust storm event using the HYSPLT cluster analysis method is shown in Figures 4.10 and 4.11. A large percentage of the PM10 trajectories analysed in the period between 10th and 30th of June 2020, reached the middle-upper troposphere of the Caribbean Sea and the Gulf of Mexico. 80%, 51%, 76%, and 70% of the PM10 particle trajectories from Bir Anzarane Morocco, Nouakchott, and Tichit Mauritania, and Bordj Badji Mokhtar Algeria arrived to the Gulf of Mexico respectively.

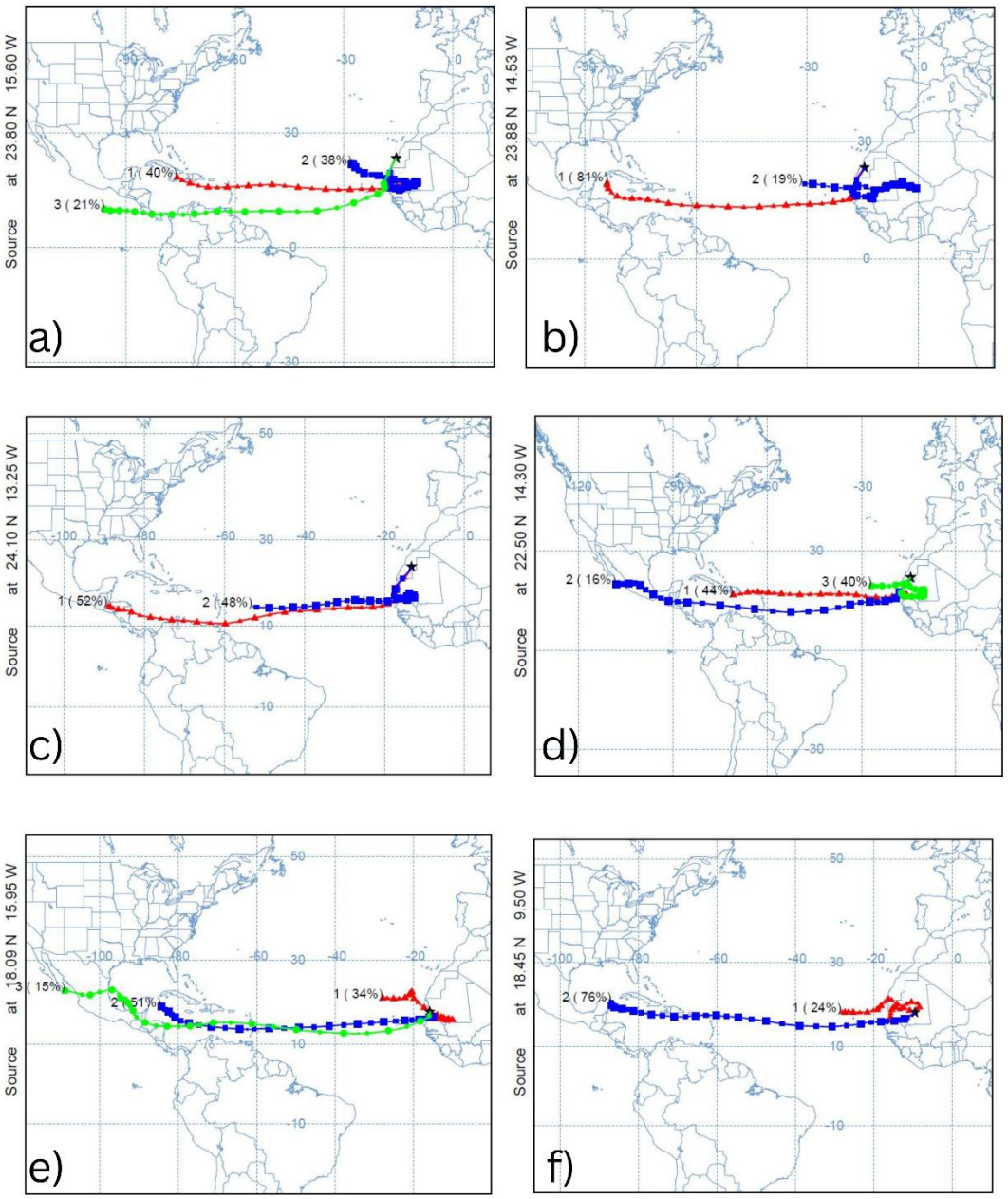

Figure 4.10. Forward trajectory cluster analysis results, each picture shows Cluster mean trajectories with the percentage of trajectories in each cluster from a) Dakhla, b) Bir Anzarane, c) Oum Dreyga, d) Aousserd, e) Nouakchott, and f) Tichit.

Many studies state that during summer, and especially during Saharan dust events, the level of PM10 and PM2.5 concentration increased dramatically. (Bozlaker *et al.*, 2013) state that during the Saharan episode in 2008, the total dust contribution for PM10 increased by 85% in Houston, Texas, which shows a dominance of the transported PM10 particles from Sahara during dust

episodes. Also, (Bozlaker *et al.*, 2019) found dust contributions of 19% to 48% of PM2.5 during the 9-day dust episode in 2014 to African dust. Additionally, the results of the cluster analysis point out a number of source regions in the western Sahara that contribute to the rise in PM10 concentrations in the Southern Coast of the United States, such as Bir Anzarane Morocco, Nouakchott, and Tichit Mauritania, and Bordj Badji Mokhtar Algeria.

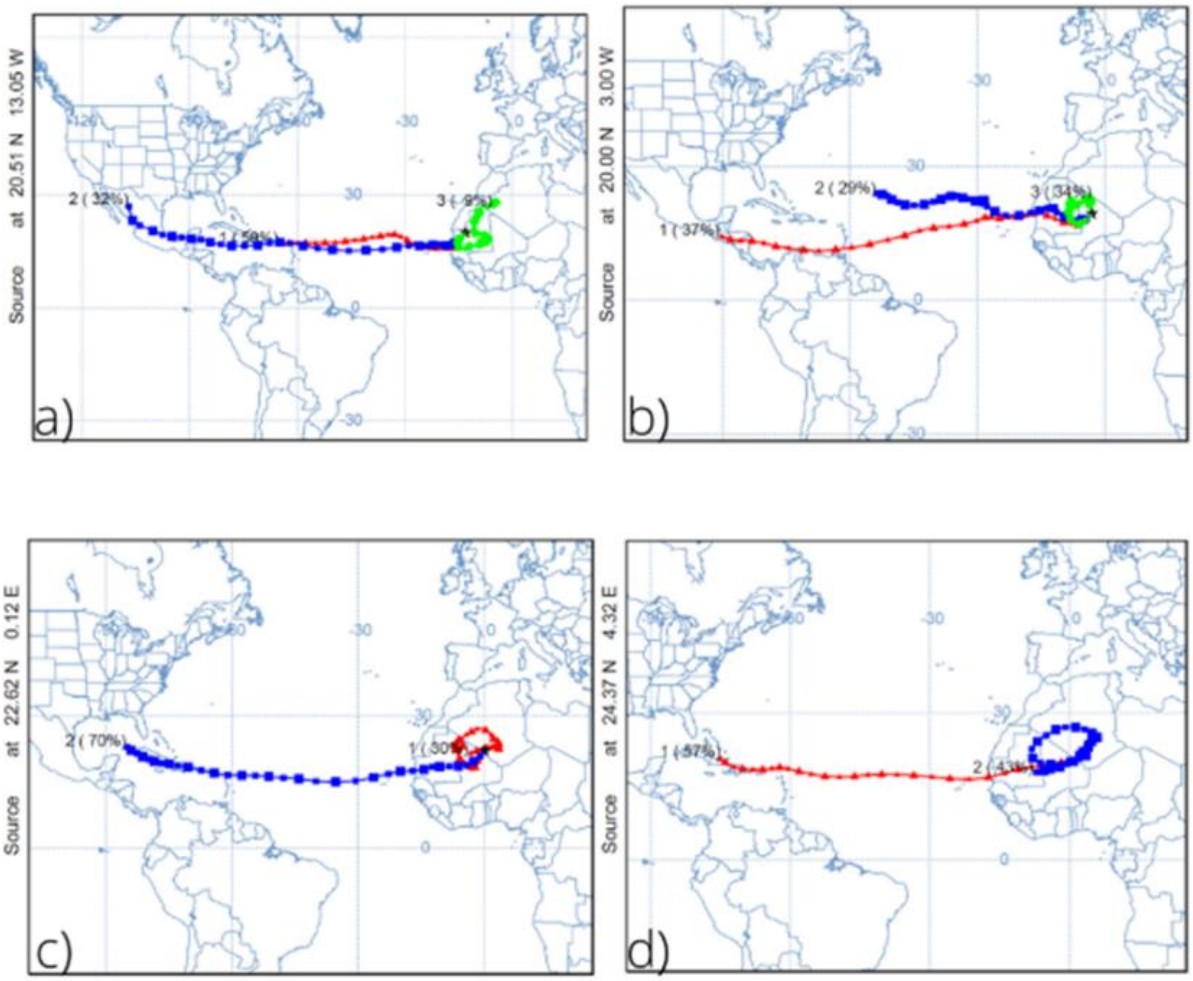

Figure 4.11. Forward trajectory cluster analysis results, each picture shows Cluster mean trajectories with the percentage of trajectories in each cluster from a) Atar, b) Toumbouctou cercle, c) Bordj Badji Mokhtar and d) Tamanrasset

4.2.1.2 PM concentration levels and AOD

In addition, the backscatter vertical profile as measured by CALIPSO on June 21 and 23, 2020 (Figure 4.12) shows evidence of the high altitude of the dust particles transported from the Saharan region. The top layer altitude of the dust on June 21 and 23 were between 4 and 4.5 km, forming a massive dust cloud (more than 2 km of thickness) over the Windward and Leeward islands in the Caribbean Sea, and the effect was seen in the hourly measurements of the PM10 concentrations of the Fort de France station in Martinique Island where the PM10 daily average concentration was 181, 264, and 183 µg/m3 on the 21, 22 and 23 of June consecutively with an hourly concentration that reached 372 µg/m3, comparing to 42 µg/m3 that was registered at the beginning of that month. Furthermore, and after 11days of the starting of the Saharan dust storm, the effect of the transported particles was clear in the US coastal cities of the Gulf of Mexico. Texas and Florida states were the most affected by having an Unhealthy level of PM10 and PM2.5 concentrations, followed by Georgia, Alabama, Mississippi, and Louisiana states that reached the level of Unhealthy for sensitive groups during the 26 and 27 June 2020, which is in correlation

with the backscatter vertical profile measured by CALIPSO on June 27, 2020, showing dust cloud over Florida state with dust top layer altitude at 4 km.

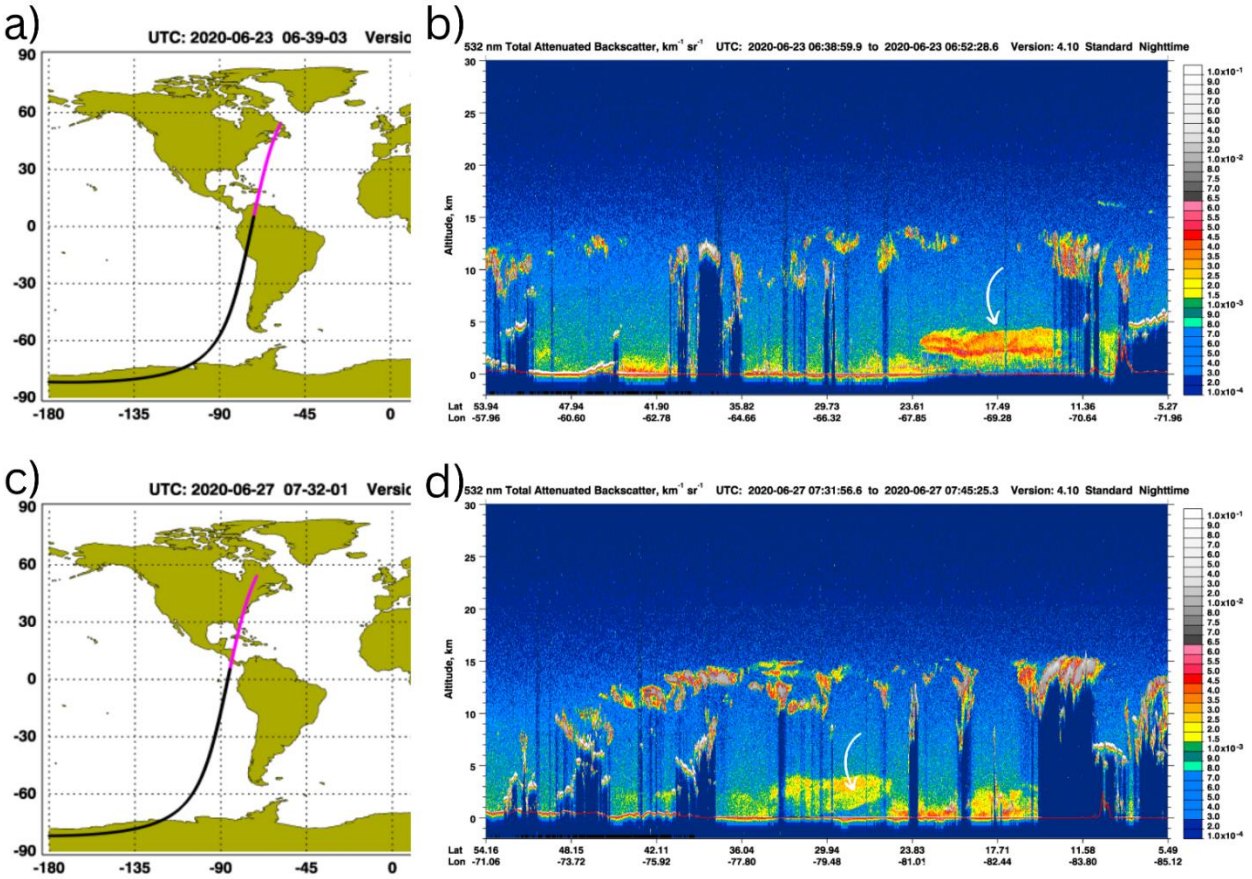

Figure 4.12. CALIPSO 532nm Total Attenuated Backscatter Version 4.10 images, the dust appears in yellow and red in the images. a) Orbit map on 23, June 2020 with area covered in d image colored in pink b) 532nm Total Attenuated Backscatter on 23, June 2020 with white arrows pointing at the dust top layer over the Caribbean Sea c) Orbit map on 27, June 2020 with area covered in image colored in pink d) 532nm Total Attenuated Backscatter on 27, June 2020 with white arrows pointing at the dust top layer over Florida state

In order to quantify the dust event, Figure 4.13 and 4.14 show AOD values retrieved from MERRA-2 re-analysis data for Bir Anzarane, Morocco and Nouakchott, Mauritania, for the month of June 2020. According to both, the June 2020 dust event was historical by June standards. For Bir Anzarane, Morocco, the highest AOD value was 3.522 in June 2020, a 188% increase from the highest value registered from 2010 to 2019 (1.87 in June 2017). For Nouakchott, Mauritania, the highest AOD value recorded in June between 2010 and 2019 was 2.78 in June 2010, while in June 2020, the highest AOD was 5.87, 211% higher. Even though such high AOD levels are

exceptional, but not uncommon; during the record-breaking March 2018 dust outbreak, Solomos *et al.* (2018) and Kaskaoutis *et al.* (2019) observed AOD values over 6.

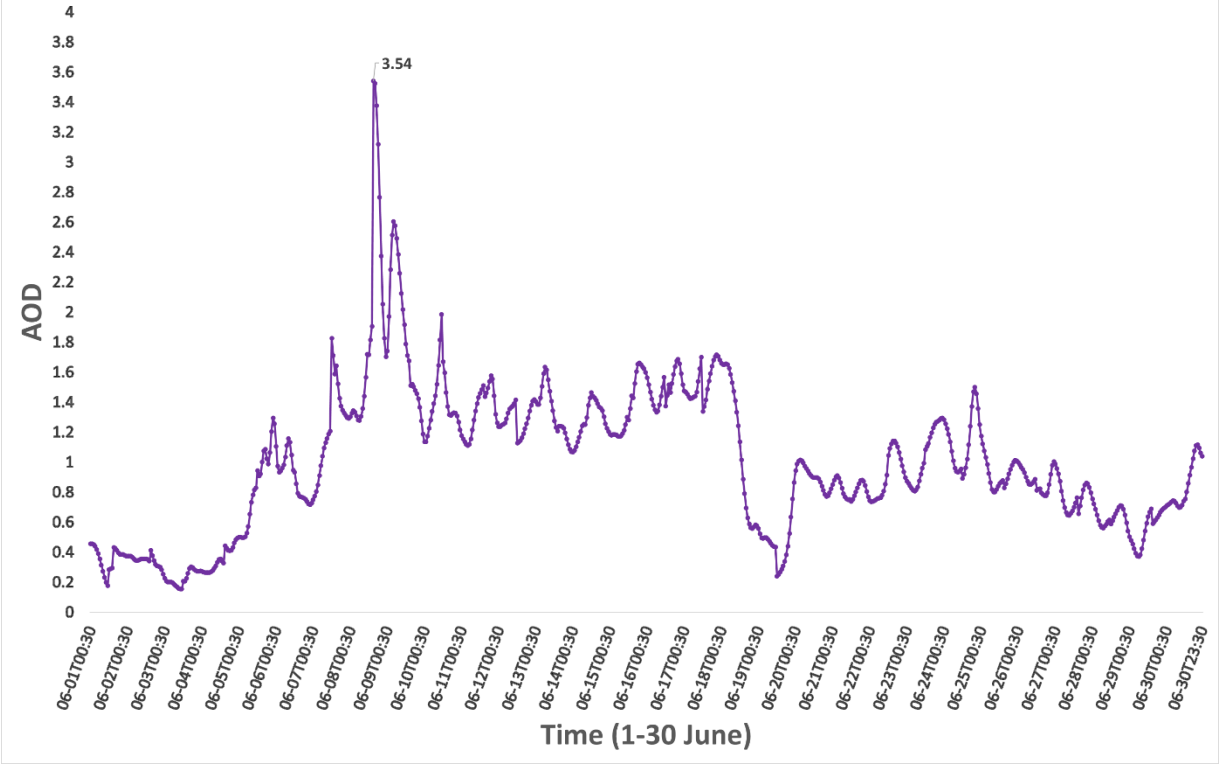

Figure 4.13. Bir Anzarane Morocco AOD values in June 2020

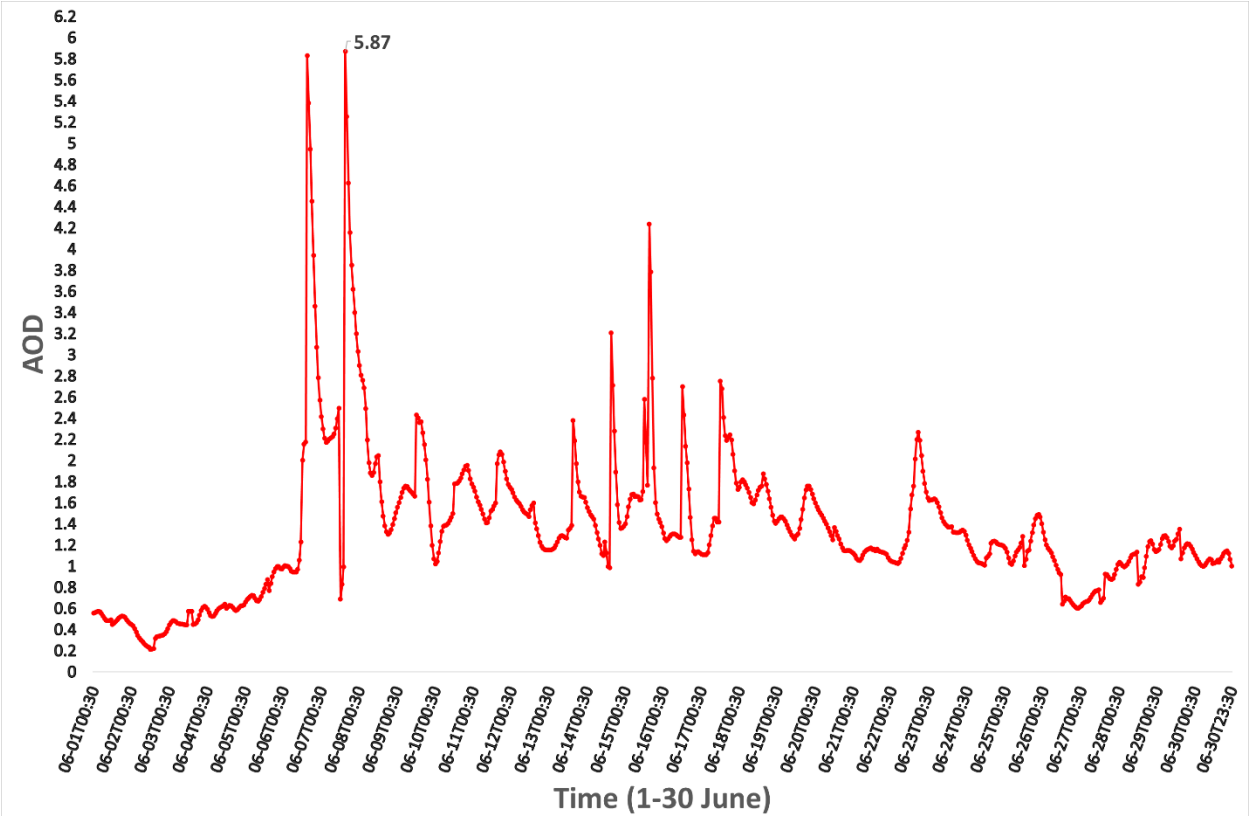

Figure 4.14. Nouakchott, Mauritania AOD values in June 2020

4.2.2 The identification and evaluation of the Saharan dust storm events in Budapest, Hungary between 2018 and 2022

Based on the daily 700 hPa geopotential height, wind maps, and SDE-specific dust transport paths, SDEs were divided into three primary synoptic meteorological groups by Varga, (2020). The various categories were distinguished by certain deterministic atmospheric patterns: Type-1 SDEs were linked to deep atmospheric depressions over Western Europe and north-western Africa. While, dust transport during Type-2 episodes was brought on by Central Mediterranean cyclones, while Type-3 events were defined based on the infrequent dust transport that occurred when dust-loaded air masses approached the Carpathian Basin from the north-western directions (Figure 4.15). From 2018 to 2022, 11 Saharan Dust events (SDEs) were identified in Hungary (Focus on the capital Budapest).

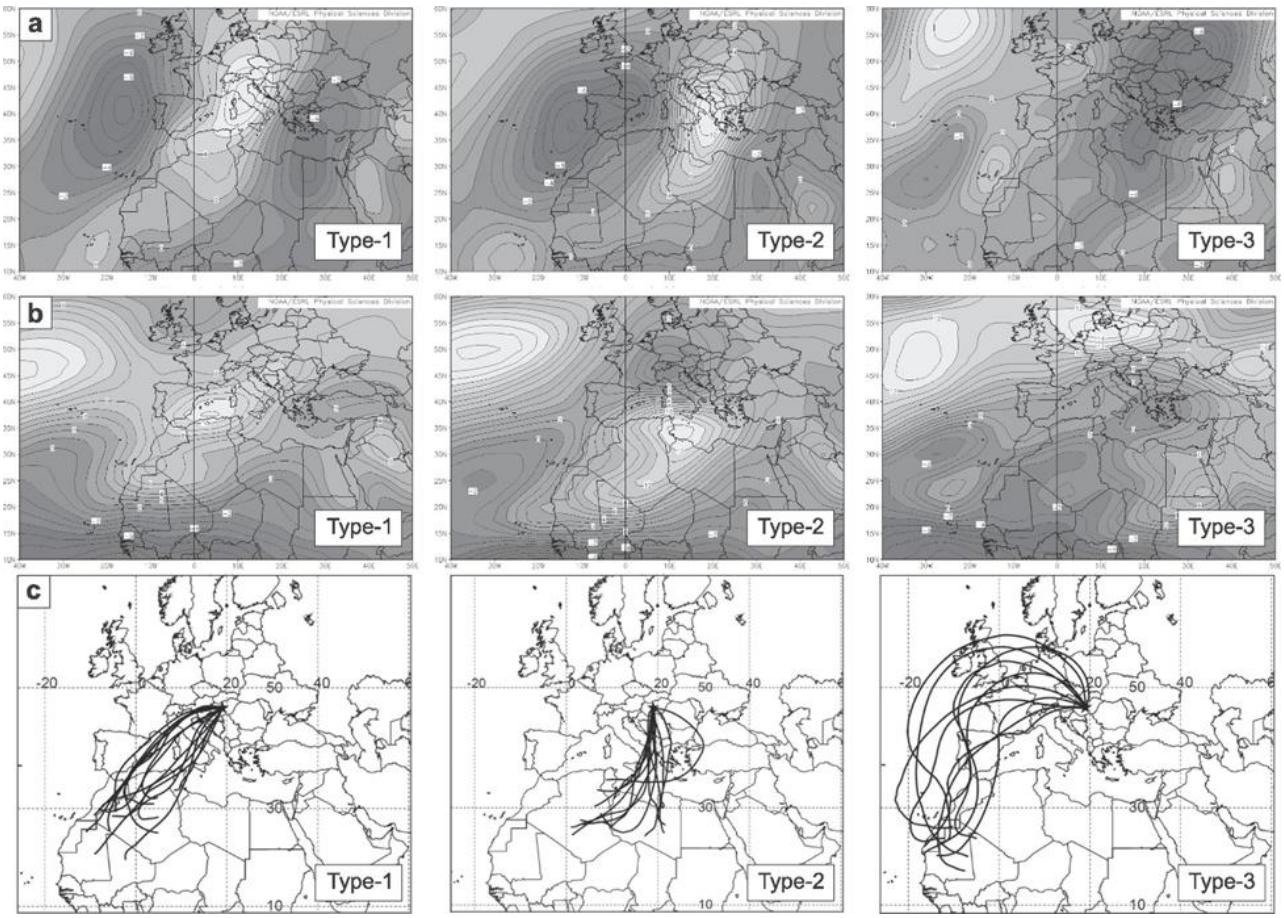

Figure 4.15. Wind flow patterns (mean meridional (a) and zonal (b) wind components at 700 hPa) and (c) specific dust transport routes at 3000 m above surface level by different Saharan dust event types (Varga, 2020).

SDE1: January 7–9, 2018

SDE1 was a type 2 event, Figure 4.16.A show the transport of the dust mass to Hungary at its peak form. The maximum Dust mass was 675.5 mg/m^2, and PM10 daily concentration increased by factor of 2.5, (from 22µg/m^3 in the 4th of January to 55 µg/m^3 in the 7th of January).

SDE2: February 7–8, 2018

SDE2 was also a type 2 as seen in Figure 4.16.B. The maximum Dust mass was 834.7 mg/m^2, and PM10 daily concentration increased by factor of 2, (from 31 µg/m^3 in the 4th of February to 66 µg/m^3 in the 9th of February).

SDE3: October 28–31/01–02 November, 2018

SDE3 was also a type 2 as seen in Figure 4.16.C, it was a two wave SDE, the first wave started to hit on the 28th of October and the second wave on the 1st of November. The maximum Dust mass was 505.5 mg/m^2 on the first wave and 367.4 mg/m^2 on the second wave, and PM10 daily concentration increased by factor of 3, (from 20 µg/m^3 in the 26th of October to 61 µg/m^3 in the 02nd of November).

SDE4: April 23-27, 2019

Even SDE4 was a type 2 as shown in Figure 4.16.D. This event was also a 2 waves SDE, the first wave started to hit on the 23rd of April and the second wave on the 26th of April. The maximum Dust mass was 993.8 mg/m^2 on the first wave and 952.9 mg/m^2 on the second wave, and PM10 daily concentration increased by factor of 1.7, (from 28 µg/m^3 in the 20th of April to 48 µg/m^3 in the 26th of April).

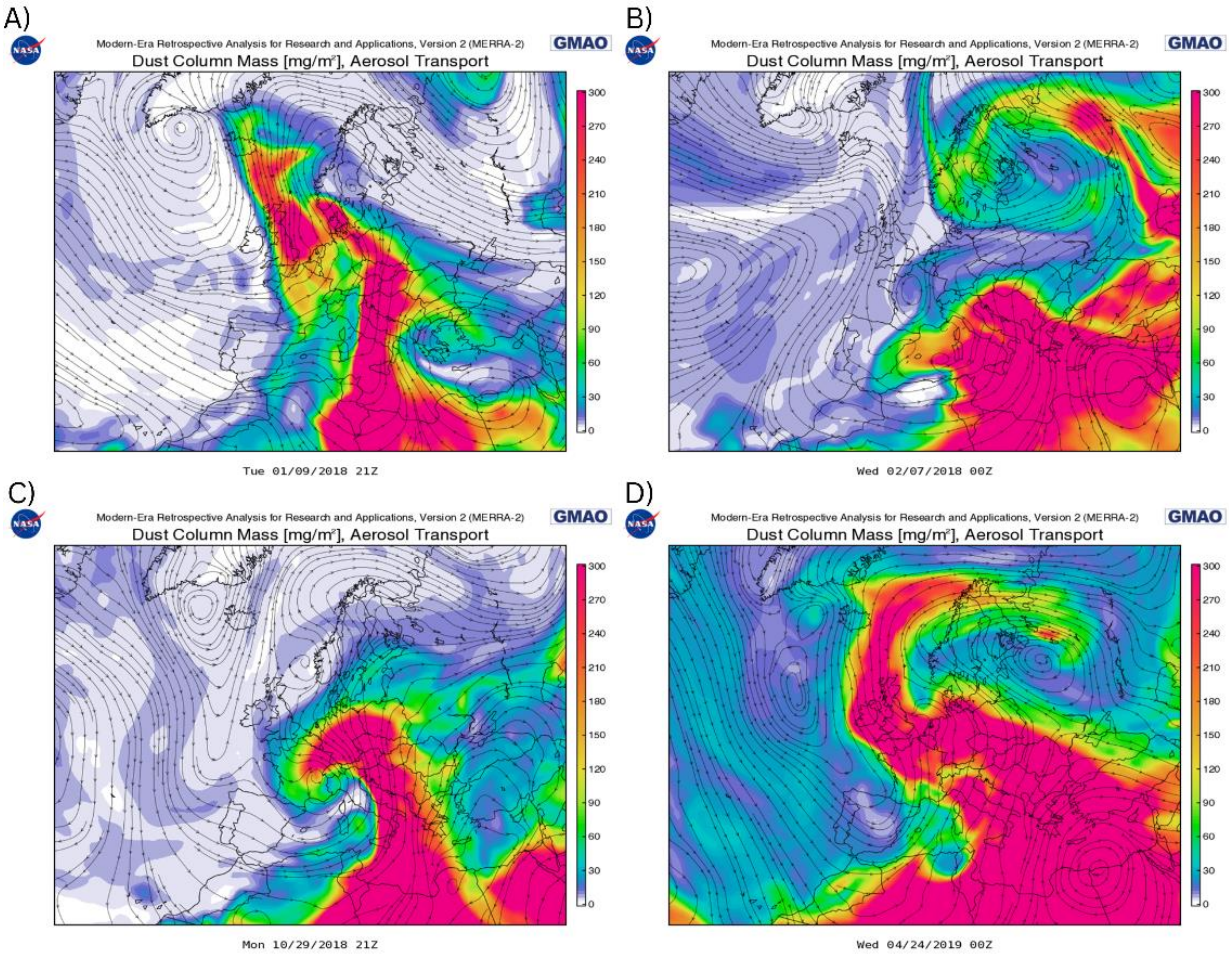

Figure 4.16. Dust Column Mass representation at its peak for A) SDE1, B) SDE2, C) SDE3, and D) SDE4.

SDE5: May 13-20, 2020

SDE5 was a type 1 as illustrated in Figure 4.17.A. This event was also a 2 waves SDE, the first wave started to hit on the 14th of April and the second wave on the 16th of April. The maximum Dust mass was 654.5 mg/m^2 on the first wave and 612.1 mg/m^2 on the second wave, and PM10 daily concentration increased by factor of 1.5, (from 18 µg/m^3 in the 12th of May to 28 µg/m^3 in the 20th of May).

SDE6: February 06-08, 2021

SDE6 is a type 1 (Figure 4.17.B). The maximum Dust mass was 989.9 mg/m^2, and PM10 daily concentration increased by factor of 2, (from 14 µg/m^3 in the 4th of February to 31 µg/m^3 in the 6th of February).

SDE7: February 23-26, 2021

SDE7 was a type 3 (Figure 4.17.C). The maximum Dust mass was 691.9 mg/m^2, and PM10 daily concentration increased by factor of 3, (from 29 µg/m^3 in the 21st of February to 92 µg/m^3 in the 26th of February).

SDE8: June 22-25, 2021

SDE8 was a type 1 (Figure 4.17.D). The maximum Dust mass was 719.8 mg/m^2, and PM10 daily concentration increased by factor of 2, (from 10 µg/m^3 in the 20th of February to 23 µg/m^3 in the 25th of February).

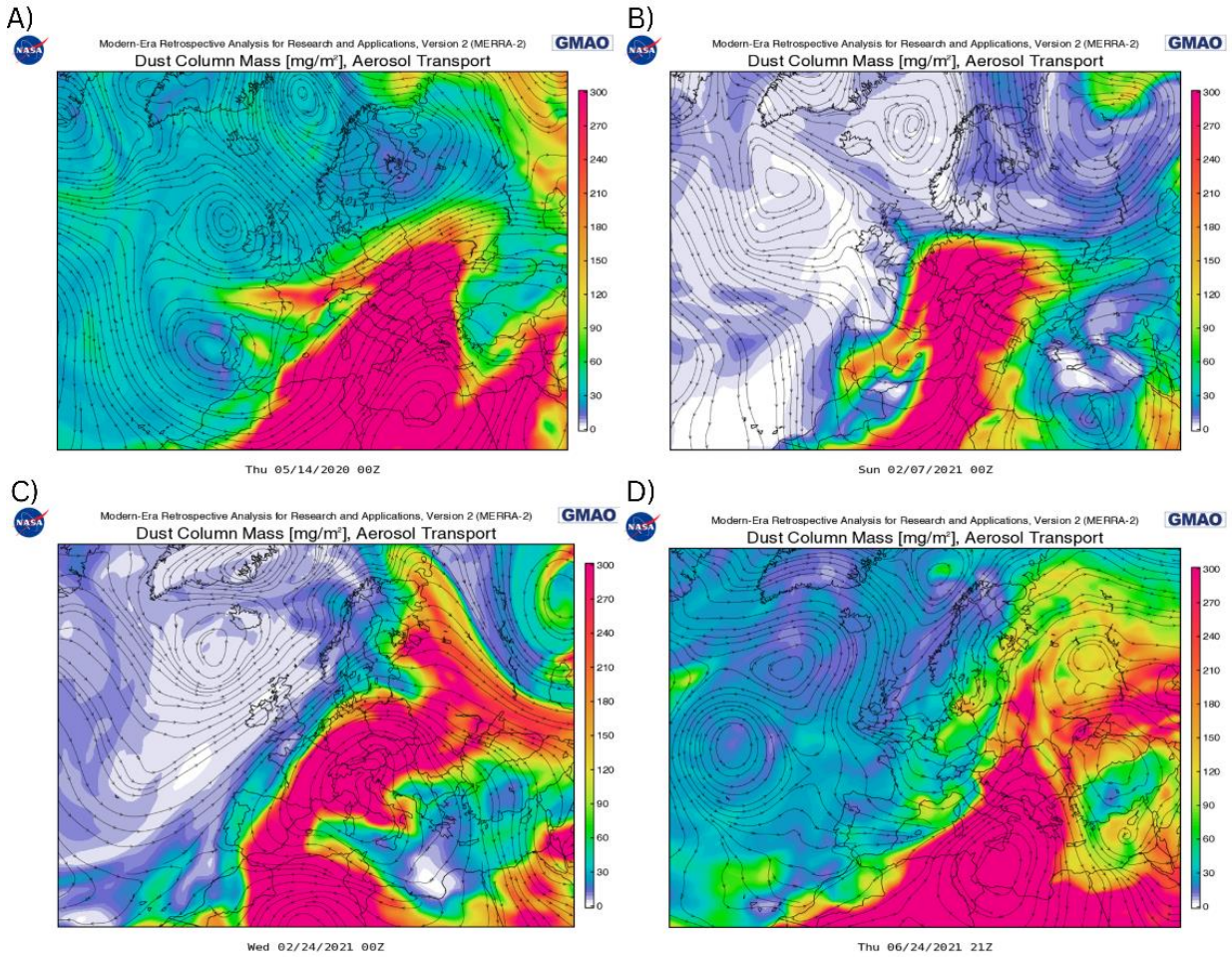

Figure 4.17. Dust Column Mass representation at its peak for A) SDE5, B) SDE6, C) SDE7, and D) SDE8.

SDE9: March 17-20, 2022

SDE9 was a type 3 (Figure 4.18.A). The maximum Dust mass was 719.8 mg/m^2, and PM10 daily concentration increased by factor of 2, (from 10 µg/m^3 in the 20th of February to 23 µg/m^3 in the 25th of February).

SDE10: March 29-31, 2022

SDE10 was a type 3 (Figure 4.18.B). The maximum Dust mass was 679.9 mg/m^2, and PM10 daily concentration increased by factor of 1.4, (from 29 µg/m^3 in the 26th of March to 43 µg/m^3 in the 30th of March).

SDE11: April 22-24, 2022

SDE11 was a type 2 (Figure 4.18.C). The maximum Dust mass was 699.7 mg/m^2, and PM10 daily concentration increased by factor of 2.4, (from 12 µg/m^3 in the 19th of April to 29 µg/m^3 in the 21st of April).

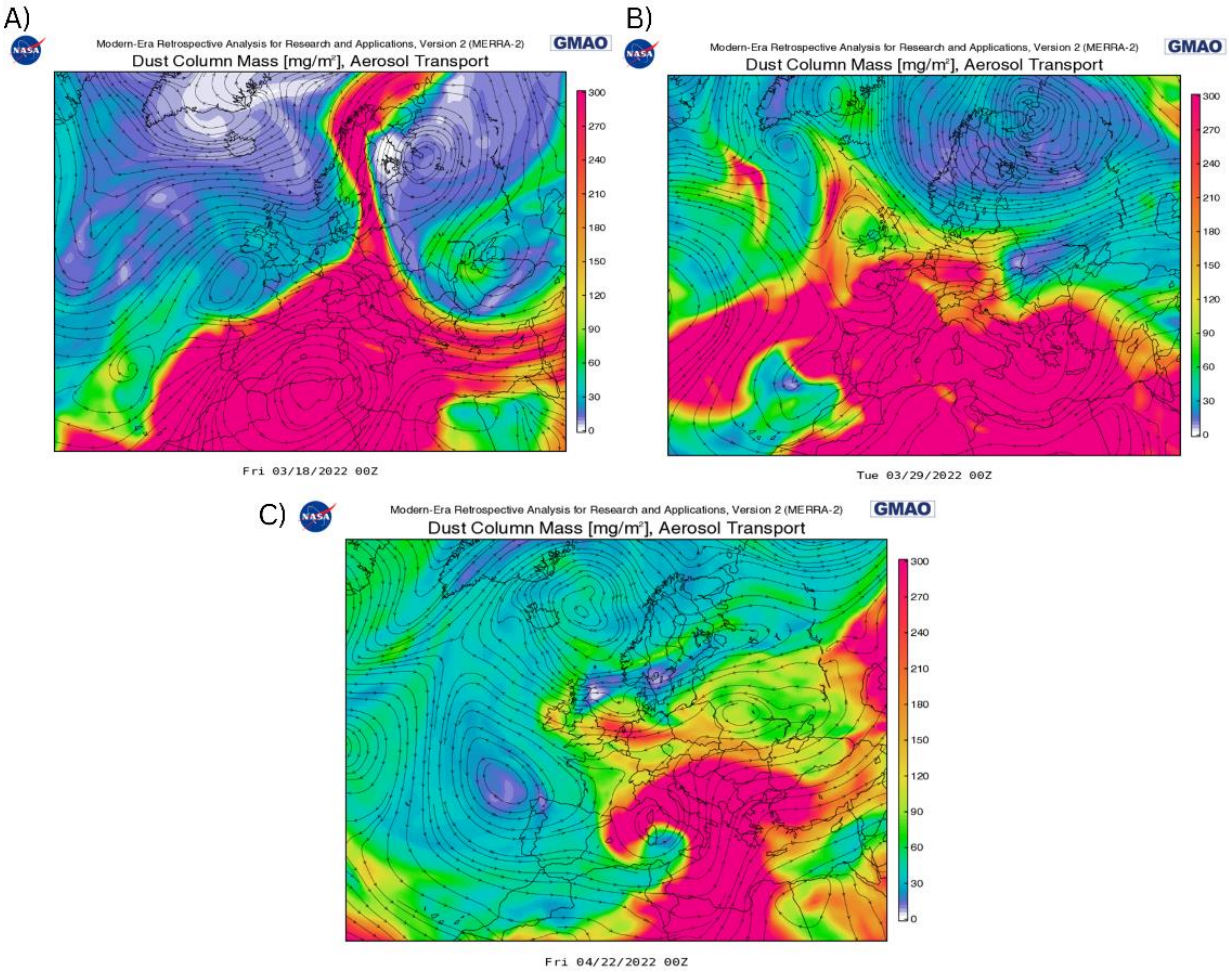

Figure 4.18. Dust Column Mass representation at its peak for A) SDE9, B) SDE10, C) SDE11

During the period of 2018 to 2022, type 2 SDEs were dominant (5 times – Figure 4.19.I), while type 1 and 3 both occurred 3 times each. In addition, February, March and April months are the months where the most of SDEs happened (7 times – Figure 4.19.II), and SDEs occurring in those months are more likely to be severe events, since the maximum dust mass registered is in that period (SDE4), and also associated with an increase of PM10 daily average concentration by a factor of 2 or more.

With increasing distance from the source, dust's grain size decreases. When transported over long distances, coarse particles typically do not exceed 20μm because of their higher settling velocity (Does *et al.*, 2016). Mahowald *et al.* (2014) hypothesized that because coarser particles tend to settle out more readily, dust in the high atmosphere is finer grained than dust that has been deposited. Moreover, a high Saharan Dust Mass during SDE could lead to high increase in PM10 concentrations, but that depends on the dust particles size and deposition velocity, which mean that a relationship between dust mass during SDE and PM10 concentration is not always a direct positive relationship. Varga, (2020), highlight contravention of the numerical simulations that estimate the mineral grains sizes during SDE, and that the bulk of global and regional dust models only use a few size-bins with a rather restricted size range, hence mineral grains larger than 20 μm are typically not taken into consideration in the numerical simulations, and the direct measurements of individual particles illustrate that the mineral grain size during a SDE in the Carpathian region is about 40 μm.

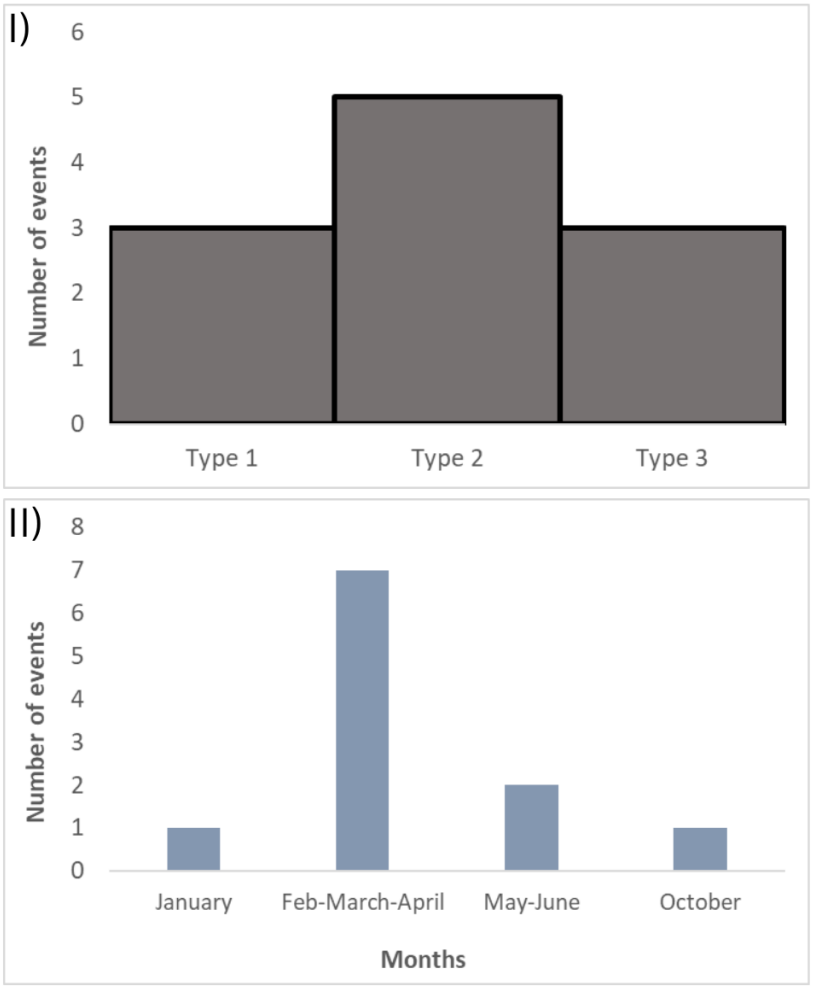

Figure 4.19. Frequency distribution of SDEs by number of events and I) Types, and II) Months of occurrence

4.2.3 Case study of the Saharan dust effects on PM10 and PM2.5 concentrations in Budapest in March 2022

4.2.3.1 PM10 and PM2.5 concentrations during the Saharan dust events

March 2022 dust storm events were type 3 events as described in previous chapter (4.4). The daily PM10 and PM2.5 concentrations increased in each Saharan dust event at a different percentage

rate. During the first SDE (SDE9) in March 2022 (17th -20th), the daily PM10 and PM2.5 concentration jumped from 34 µg/m^3 and 18 µg/m^3 in 16th to 46 µg/m^3 and 28 µg/m^3 in the 18th and then start to decrease to reach 25 µg/m^3 and 15 µg/m^3 in the 20th. For the second SDE (SDE10) in March 2022 (28th – 31st), the daily PM10 and PM2.5 concentration changed from 29 µg/m^3 and 13 µg/m^3 on the 27th to 43 µg/m^3 and 18 µg/m^3 on the 30th after which begin to decline to attain 16 µg/m^3 and 14 µg/m^3 in the 1st of April 2022. Hourly PM10 and PM2.5 concentrations (Figure 4.20) provide details on how the hourly concentration changed with the SDE.

SDE9 was more intense than SDE10, as the effects were seen on the level of PM10 and PM2.5. The peak hourly concentration for PM10 was 86 µg/m^3 and 57 µg/m^3 for SDE9 and SDE10 respectively, while for PM2.5 it reached 51 µg/m^3 and 27 µg/m^3 as hourly concentration for SDE9 and SDE10 respectively.

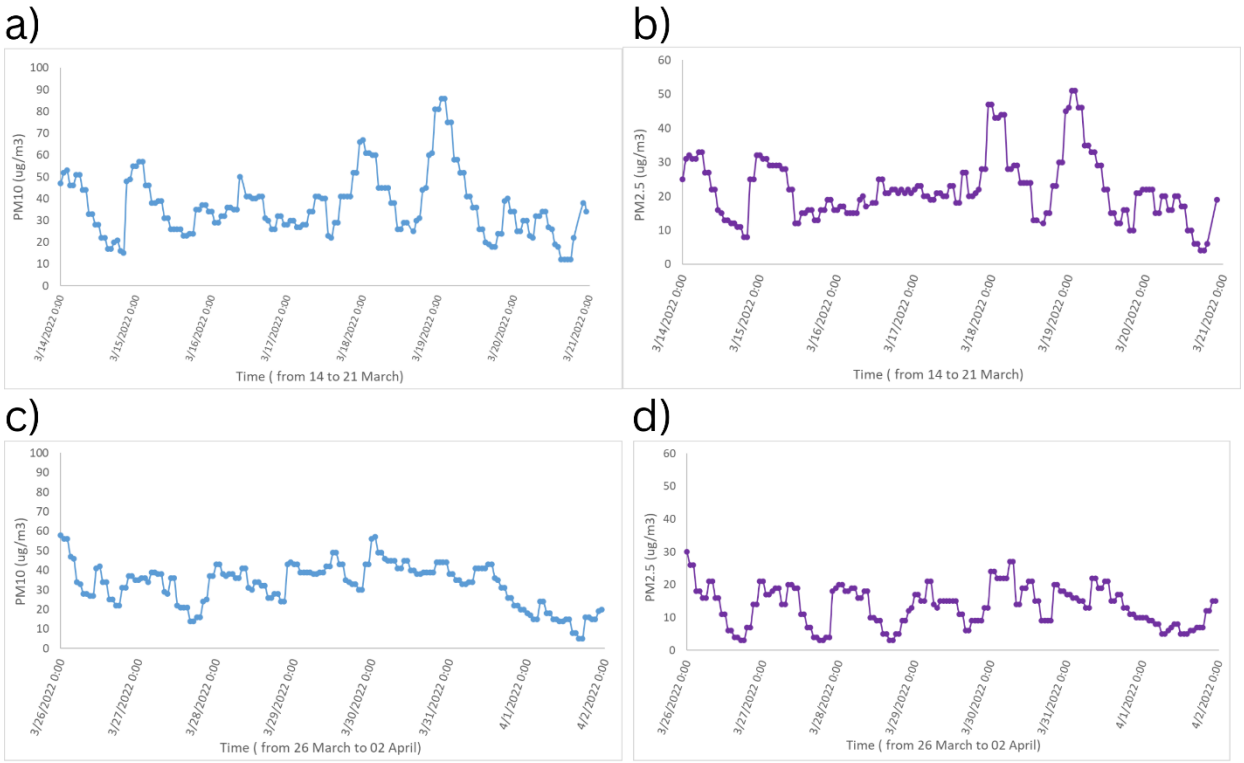

Figure 4.20. Hourly concentration of a) PM10 during SDE9, b) PM2.5 during SDE9, c) PM10 during SDE10, and d) PM2.5 during SDE10

About weather conditions, from 8 to 12 March 2022, the daily maximum temperature was between 7 and 10˚C, the wind direction from the North and Northeast direction, and the maximum wind speed was between 3.5 and 5 m/s, and no precipitations were registered. There was a slight increase in the maximum temperature (between 8 and 15˚C) in the following days, while from the 15th of March wind pattern full of Saharan dust arrived by western winds and in-ground surface the wind speed didn't exceed 3.5 m/s with no precipitations during SDE9.

The temperature continued to increase after SDE9, ranging from 16 and 21˚C as the daily maximum temperature, and start to decrease from the 31st of march and returns to the same levels that were at the beginning of March 2022, from the 2nd of April 2022 (between 5 and 9˚C). On the other hand, Saharan dust clouds were transported by West-Southwest, Southwest, and South-Southwestern winds, and maximum wind speed on the ground surface ranged from 3 and 5 m/s during the SDE10, and start to increase from the 1st of April 2022 to exceed 6 m/s as maximum wind speed, and no precipitations occurred on those days.

With increasing distance from the source, dust's grain size decreases. When transported over long distances, coarse particles typically do not exceed 20 μm because of their higher settling velocity. (Mahowald *et al.*, 2014; Does *et al.*, 2016) thought that dust in the upper atmosphere is finer grained than dust that has been deposited because coarser particles drop off more rapidly. Seasonally, summertime is when Saharan dust is coarser than wintertime. The high dust load in both SDEs triggered the increase of hourly PM10 and PM2.5 concentration.

The Sahara is the main source of dust in the Northern Hemisphere, and it is obvious that it has an impact on many different continents, from the fertilization of South America to the air quality in Europe. The Saharan dust storms affects the PM concentrations depending on the intensity of the storm. During March 2022 SDEs, PM10 and PM2.5 concentrations in an urban background air quality station in Budapest increased by 12 μg/m^3 and 10 μg/m^3 respectively during SDE9, and 14 μg/m^3 and 5 μg/m^3 during SDE10 as daily average concentrations. in Both SDEs the effect on PM10 was almost the same, while SDE9 raised the PM2.5 concentrations more than SDE10.

The effects of March 2022 Saharan dust events were similar to the one that was identified in 2016, on October 15th, which washed out a large amount of Saharan dust in the central European region, where it has several impacts, from flight security to air quality and impacts on solar radiation (Rostási *et al.*, 2022). However, March 2022 Saharan dust events had a more significant impact on air quality all over Europe, affecting a wider geographical area of western, central, and northern Europe, from Spain up to Scandinavia that led to an increase in PM concentrations (Liaskoni *et al.*, 2023; Uzunpinar *et al.*, 2023), and was associated with dust-infused cirrus clouds that persisted for nearly a week, affecting weather patterns and cloud cover over the region.

With the changing of the world climate, the intensity and the number of the Saharan dust storms episodes increase, and many models are still improving to provide more accurate forecast and to analyse the dust effects on different meteorological and air quality parameters.

4.3 Estimation and evaluation of PM concentrations

4.3.1 Evaluation of PM surface concentrations simulated by Version 5.12.4 of NASA's MERRA-2 Aerosol Reanalysis over Hungary in the period between 2019 and 2021

4.3.1.1 First Approach

In the machine learning algorithm, we used a split of 0.8x0.2. 80% of the data were used to train the model and 20% for the validation of the predicted values given by the trained model. To arrive to the results presented, we tried the model for many times, and each time the parameters of the machine learning algorithms (mainly the number of trees) were changed until arriving to the maximum results that can be achieved, where beyond that point the results whether they stopped improving or the performances start to decline.

For all the location chosen for this study, the use of equations 1 and 2 to estimate PM10 and PM2.5 result in an R^2 less than 0.1, and R^2 improved when coupling 5 species concentration used in equations 1 and 2 with meteorological data and AOD (Figure 4.21).

For Budapest Gilice tér station, in case of estimating PM10, MLR and OLS had low and a similar R^2 (0.22 and 0.21), RMSE (15.4 and 15.5 μg/m^3) and MAE (11.4 and 11.5 μg/m^3) values. For RF, R^2 of 0.75 achieved when N (the number of trees) was equal to 300, and RMSE and MAE were 9.1 μg/m^3 and 6.4 μg/m^3 respectively. While, for XT R^2 was equal to 0.78 when N=300, and RMSE and MAE were 8.1 μg/m^3 and 5.7 μg/m^3 respectively. Additionally, in case of estimating PM2.5, MLR and OLS also had low and a similar R^2 (0.27), RMSE (9.1 μg/m^3) and MAE (6.7 μg/m^3) values. For RF maximum value of R^2 (0.75) obtained when N was equal to 300, RMSE and MAE were 5.3 μg/m^3 and 3.5 μg/m^3 respectively. While, for XT, R^2 was equal to 0.8 when N=300, RMSE and MAE were 4.7 μg/m^3 and 3.1 μg/m^3 respectively.

The same was for Kecskemét, for PM10, low R^2 of 0.17 and 0.16 were obtained in MLR and OLS model respectively, and RMSE were 15.5 µg/m^3 and 15.6 µg/m^3, and MAE was identical for both models (11.5 µg/m^3). For RF, value of 0.66 for R^2 when N=270, and RMSE and MAE were 9.8 µg/m^3 and 6.8 µg/m^3 respectively. While, for XT R^2 was equal to 0.75 when N=300, and RMSE and MAE were 8.5 µg/m^3 and 6 µg/m^3 respectively. Moreover, in case of estimating PM2.5, MLR and OLS again had low R^2 (0.23 and 0.22 respectively), RMSE (11.3 µg/m^3 for both models) and MAE (7.5 µg/m^3 for both models). For RF maximum value of R^2 (0.69) obtained when N was equal to 300, RMSE and MAE were 7.1 µg/m^3 and 4.7 µg/m^3 respectively. While, for XT, R^2 was equal to 0.76 when N=300, RMSE and MAE were 6.2 µg/m^3 and 3.9 µg/m^3 respectively.

Furthermore, for Kazincbarcika in case of PM10, R^2 of 0.24 was obtained in MLR and OLS models, and RMSE was 18.4 µg/m^3 and 18.5 µg/m^3, and MAE 13.3 µg/m^3 and13.4 µg/m^3 respectively. For RF, value of 0.7 for R^2 when N=300, and RMSE and MAE were 11.4 µg/m^3 and 7.6 µg/m^3 respectively. While, for XT, R^2 was equal to 0.77 when N=300, and RMSE and MAE were 10 µg/m^3 and 6.6 µg/m^3 respectively. Additionally, in case of PM2.5, MLR and OLS again had low R^2 (0.34 and 0.33 respectively), RMSE (15.3 µg/m^3 for both models) and MAE (11.2 µg/m^3 for both models). For RF maximum value of R^2 (0.75) obtained when N was equal to 265, RMSE and MAE were 9.4 µg/m^3 and 6.3 µg/m^3 respectively. While, for XT, R^2 was equal to 0.8 when N=300, RMSE and MAE were 8.3 µg/m^3 and 5.5 µg/m^3 respectively.

a)

PM10	R2	RMSE	MAE
Eq1	>0.1	22.7	15.7
MLR	0.22	15.4	11.4
RF	0.72	9.1	6.4
XT	0.78	8.2	5.7
OLS	0.21	15.5	11.5

b)

PM2.5	R2	RMSE	MAE
Eq2	>0.1	11.7	7.9
MLR	0.27	9.1	6.7
RF	0.75	5.3	3.5
XT	0.8	4.7	3.1
OLS	0.27	9.1	6.7

c)

PM10	R2	RMSE	MAE
Eq1	>0.1	24.9	15.1
MLR	0.17	15.5	11.5
RF	0.66	9.8	6.8
XT	0.75	8.5	6
OLS	0.16	15.6	11.5

d)

PM2.5	R2	RMSE	MAE
Eq2	>0.1	14.2	8.9
MLR	0.23	11.3	7.5
RF	0.69	7.1	4.4
XT	0.76	6.2	3.9
OLS	0.22	11.3	7.5

e)

PM10	R2	RMSE	MAE
Eq1	>0.1	26	17
MLR	0.24	18.4	13.3
RF	0.7	11.4	7.6
XT	0.77	10	6.6
OLS	0.24	18.5	13.4

f)

PM2.5	R2	RMSE	MAE
Eq2	>0.1	22.1	14.2
MLR	0.34	15.3	11.2
RF	0.75	9.4	6.3
XT	0.8	8.3	5.5
OLS	0.33	15.3	11.2

Figure 4.21. Tables of performance statistic parameters for Budapest Gilice tér a) PM10, b) PM2.5, and Kecskemét c) PM10, d) PM2.5, and Kazincbarcika e) PM10 and f) PM2.5

The MERRAero dataset is useful tool to estimate PM10 and PM2.5 concentrations. The results show that the congruence in hourly PM10 and PM2.5 values between the observation and the calculated values based on equations 1 and 2 was inconsequential in all locations of the study. However, estimated PM10 and PM2.5 got better when coupling the estimations with meteorological data and component concentrations used in equations 1 and 2. In the three locations chosen for this study, MLR and OLS had poor R^2 (between 0.16 and 0.34), while the best R^2 was always achieved in case of XT model. The high RMSE and MAE results in case of Kazincbarcika compared to Budapest and Kecskemét, is due to the fact that concentrations of PM10 and PM2.5 registered in Kazincbarcika are higher than in Budapest and Kecskemét.

The use of sophisticated machine learning algorithms like RF and XT, gave better estimations of PM10 and PM2.5, in comparison to linear regression machine learning (MLR and OLS), and that is because of the complicated non-linear relationship between PM10 and PM2.5 to other variables like meteorological data.

4.3.1.2 Second Approach

Figure 3.7, describe the method used in this section. The estimation in this approach was done only using 4 machine learning algorithms (MLR, OLS, RF, and XT) to estimate PM2.5 concentrations based on 10 variables (AOD, O_3, NO_2, SO_2, T, Ws_{10}, Ws_{50}, RH, P, and PBLH). Figure 4.22 summarize the results of the MLR, OLS, RF, and XT models.

Multiple linear regression and Ordinary least square regression:

Except for Kazincbarcika where R^2 was in good range (0.65 and 0.64) for MLR and OLS, the value obtained in case of Budapest Gilice tér and Kecskemét were low values (between 0.29 and 0.32).

Random Forest:

For RF, the peak performance was achieved when N=1700 for the three locations. The results show that R^2 value using the RF regression machine learning algorithm was 0.69, 0.71, 0.83 for Budapest Gilice tér, Kecskemét, and Kazincbarcika respectively, overall RMSE was 5.9, 6.9 and 7.9 $\mu g/m^3$ and MAE was 4, 4.5 and 5.1 $\mu g/m^3$ respectively.

Extra Tree regression:

For XT, the peak performance was achieved when N =1000 for Kecskemét, and Kazincbarcika and 1100 for Budapest Gilice tér. The results show that R^2 value between the estimated and observed PM2.5 using the XT regression machine learning algorithm was 0.73, 0.75, and 0.84 for Budapest Gilice tér, Kecskemét, and Kazincbarcika respectively, overall RMSE was 5.5, 6.4, and 7.6 $\mu g/m^3$, and MAE was 3.7, 4.2, and 4.8 $\mu g/m^3$.

I)	R2	RMSE	MAE
MLR	0.3	8.9	6.5
RF	0.69	5.9	4
XT	0.73	5.5	3.7
OLS	0.29	9	6.5

II)	R2	RMSE	MAE
MLR	0.32	11.2	7.5
RF	0.71	6.9	4.5
XT	0.75	6.4	4.2
OLS	0.32	11.2	7.5

III)	R2	RMSE	MAE
MLR	0.65	11.4	8.1
RF	0.83	7.8	5.1
XT	0.84	7.6	4.8
OLS	0.64	11.5	8.2

Figure 4.22. Tables of performance statistic parameters for I) Budapest Gilice tér, II) Kecskemét, and III) Kazincbarcika

In RF algorithm, each tree in the ensemble is constructed from a sample selected with substitute from the training set. Additionally, while partitioning each node throughout tree construction, the optimum split is determined by selecting either all input features or a random subset of size. The goal of these two randomness sources is to reduce the variance of the forest estimator. Individual decision trees, in fact, have a large variation and tend to overfit. Forests with injected randomness provide decision trees with partially dissociated prediction errors. Some inaccuracies can be eliminated by taking an average of such projections. RFs minimize variance by merging various trees, sometimes at the expense of a modest bias increase. In reality, the variance decrease is frequently large, resulting in a superior overall model. The way splits are produced in XT algorithm goes even further. A random subset of candidate features is employed, much as in RF, but instead of looking for the most discriminative thresholds, thresholds are produced at random for each candidate feature, and the best of these randomly-generated thresholds is chosen as the splitting criterion. This generally allows for a little reduction in model variance at the price of a slight increase in bias.

The MERRAero dataset is a valuable tool for estimating PM10 and PM2.5 concentrations. The results show that the congruence in hourly PM10 and PM2.5 values between the observation and the calculated values based on equations 1 and 2 was inconsequential in all study locations. However, estimated PM10 and PM2.5 improved when coupling the estimations with

meteorological data and component concentrations used in equations 1 and 2. In the three sites chosen for this study, MLR and OLS had poor R^2 (between 0.16 and 0.34), while the best R2 was always achieved in the case of the XT model. The high RMSE and MAE results in the case of Kazincbarcika compared to Budapest and Kecskemét can be noticed due to the concentrations of PM10 and PM2.5 registered in Kazincbarcika, which are higher than in Budapest and Kecskemét.

In addition to the absence of nitrate particle concentrations, Provençal *et al.* (2017) explains the incongruence between observed and simulated PM2.5, which is probably due to a combination of [SO_4], [OC] and [BC] differences. Additionally, Buchard *et al.* (2016) noticed a disparity in carbonaceous particle concentrations in suburban areas of the United States. Many additional research has proposed adding nitrate concentrations to improve MERRA-2 PM2.5 estimates (He *et al.*, 2019; Ma, Xu and Qu, 2020), while poor MERRA-2 PM estimations were claimed to be caused mostly by the use of the Goddard Earth Observing System, version-5 (GEOS-5) model's bottom-up emission database and meteorological issues in GOES-5 simulations (Song *et al.*, 2018). According to Ali *et al.* (2022), multiple statistical models can be used to estimate PM2.5 using MERRA-2 aerosol reanalysis data, with the random forest model having the highest accuracy. Their results indicate that the random forest model is an appropriate choice for calculating PM2.5 concentrations in China.

It was demonstrated that machine learning is a valuable method for predicting PM2.5 by using algorithms to estimate PM2.5 based on MERRA-2 AOD, Meteorological, NO_2, O_3, and SO_2 data in 3 years (2019 to 2021) for Budapest. A comparison between 4 machine learning approaches revealed that the Extra-Tree regression model outperformed other models like RF, MLR and OLS. For Budapest the results of XT model for estimation of PM2.5 give an R^2 of 0.73, RMSE of 5.5 μg/m^3, and MAE of 3.7 μg/m^3. For Kecskemét an R^2 of 0.75, RMSE of 6.4 μg/m^3, and MAE of 4.2 μg/m^3. And for Kazincbarcika an R^2 of 0.84, RMSE of 7.6 μg/m^3, and MAE of 4.8 μg/m^3. The use of sophisticated machine learning algorithms like RF and XT gave better estimations of PM10 and PM2.5, compared to linear regression machine learning (MLR and OLS); that is because of the complicated non-linear relationship between PM10 and PM2.5 to other variables like meteorological data.

4.3.2 Calibration of CAMS PM2.5 data over Hungary using machine learning

Python 3.9.17 was used to write a code that performs data preprocessing, model training, prediction, evaluation, and visualization, for the data using the LightGBM regression model.

Promising results were obtained from the calibration of CAMS PM2.5 data using the LightGBM algorithm. The correlations before and after training the model were analysed, revealing noticeable improvements in prediction accuracy (Figures 4.23 and 4.24). Before training, the correlations between the observed and CAMS PM2.5 data varied across the stations, ranging from 0.07 to 0.20. However, after training, the correlations significantly increased, ranging from 0.78 to 0.88. These enhanced correlations demonstrate the efficacy of the LightGBM algorithm in capturing the relationships between the input features and PM2.5 levels, leading to improved accuracy in predicting air quality.

The evaluation metrics, such as the R^2 scores and root mean squared error (RMSE), were utilized to assess the model's performance. The R^2 scores, which measure the model's ability to explain the variance in observed PM2.5 values, ranged from 0.61 to 0.77. This indicates that the model accounted for 61% to 77.4% of the variance, indicating a good fit to the data. Furthermore, the RMSE values, representing the average magnitude of the differences between predicted and observed values, ranged from 5.31 to 9.92 μg/m^3. Lower RMSE values indicate higher precision and accuracy in the model's predictions.

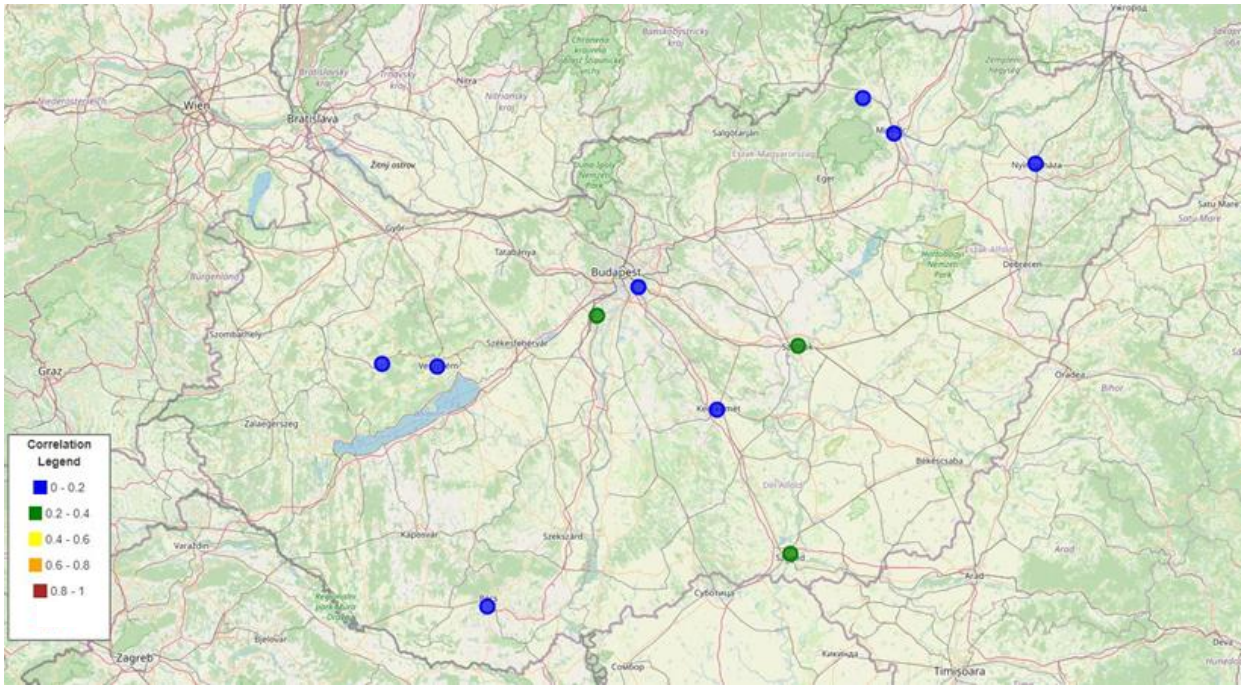

Figure 4.23. Correlation map between CAMS and In-situ PM2.5 before calibration

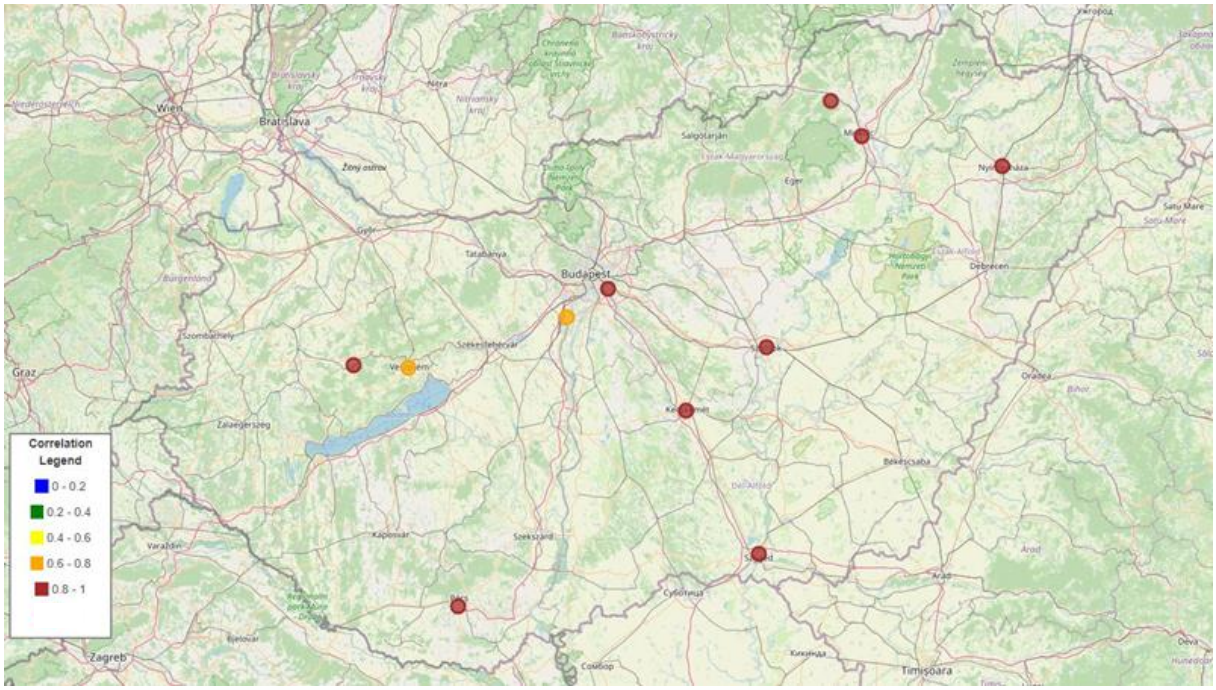

Figure 4.24. Correlation map between the predicted and In-situ PM2.5 after calibration

Furthermore, the scatter plots (Figures 4.25 and 4.26) also shed light on the model's predictive power by showcasing the proximity of the predicted PM2.5 values to the observed values. The Figures 4.25 and 4.26 show scatter plots of four stations and the rest of the stations can be found in appendices section. The close alignment between the predicted and in situ PM2.5 data points in the scatter plots signifies the model's ability to capture the underlying patterns and relationships. The proximity between these points reinforces the improved correlations observed after training, substantiating the effectiveness of the LightGBM algorithm in calibrating CAMS PM2.5 data. The plots demonstrate enhanced correlations and close alignment between predicted and in situ PM2.5 data points, highlighting the algorithm's ability to accurately predict PM2.5 levels.

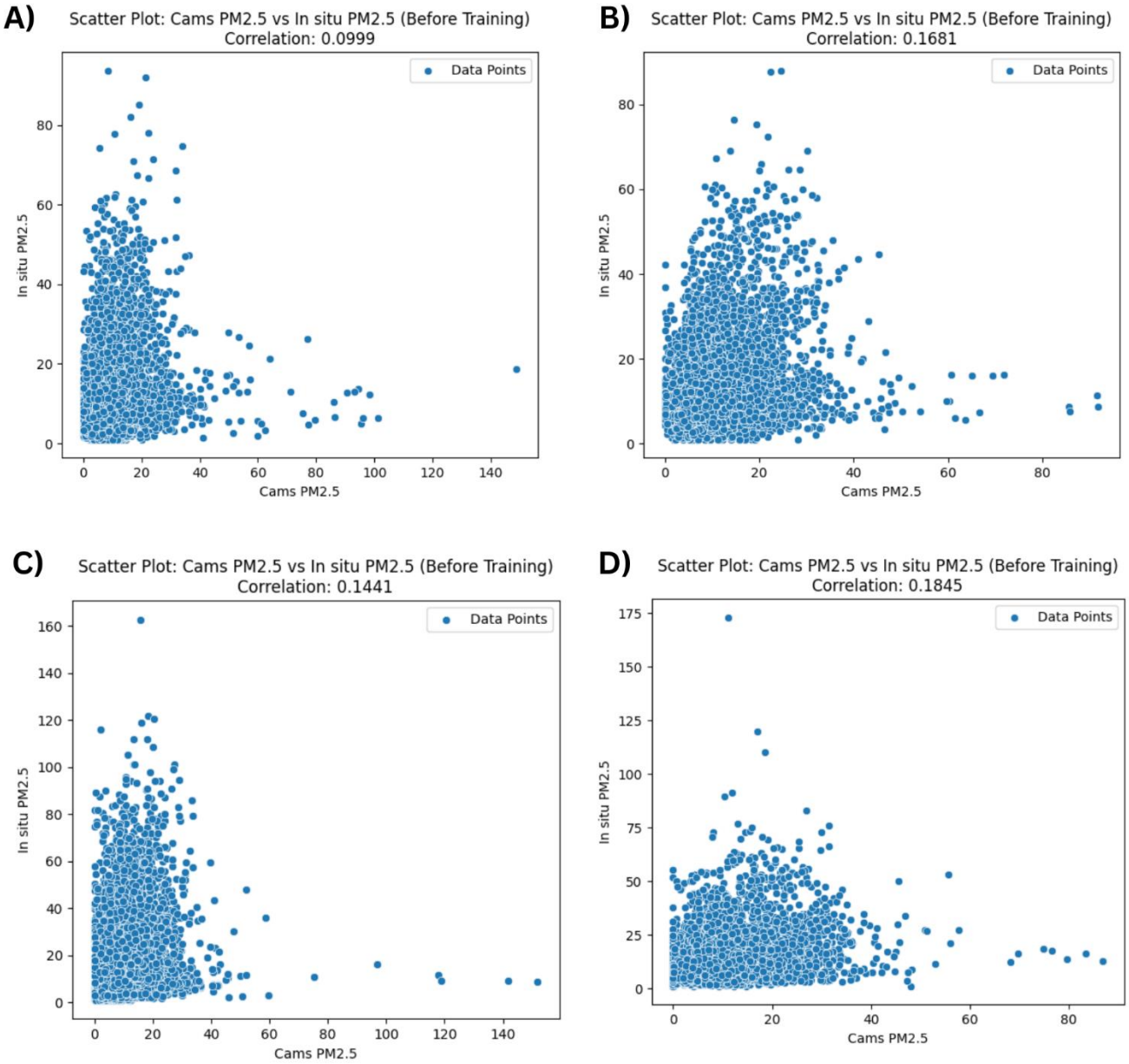

Figure 4.25. Scatter plots of raw CAMS PM2.5 data for A) Ajka station, B) Budapest Gilice ter station, C) Kazincbarcika station, D) Kecskemet station

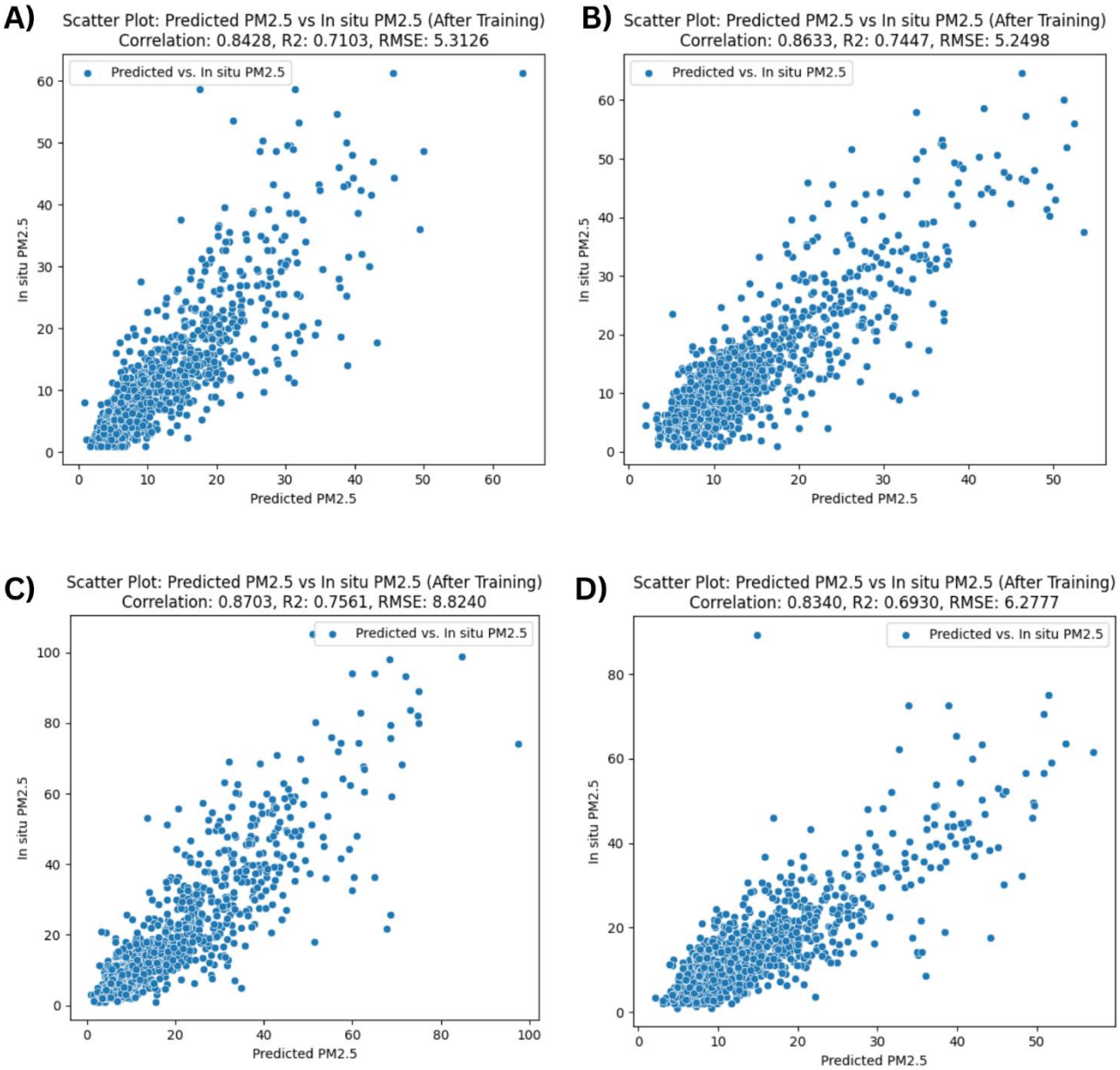

Figure 4.26. Scatter plots of predicted and In-situ PM2.5 for A) Ajka station, B) Budapest Gilice ter station, C) Kazincbarcika station, D) Kecskemet station

The results of this study calibration of CAMS PM2.5 data using the LightGBM algorithm align with other studies' conclusions, emphasizing the necessity of increasing the accuracy and dependability of CAMS datasets. In agreement with Ali *et al.* (2022), who reported CAMS overestimation in several places, our findings suggest that CAMS PM2.5 data had greater correlation values after calibration. Before training, the correlation values ranged from 0.0719 to 0.2072, demonstrating a moderate relationship between CAMS PM2.5 and in-situ PM2.5 readings. However, following training, the correlation values improved significantly, ranging from 0.7869 to 0.8820, indicating a greater link between the calibrated PM2.5 levels and in-situ observations. (Gueymard and Yang, 2020) also emphasized the limits of raw CAMS PM2.5 estimates, such as coarse spatial resolution and modelling biases. Our findings back with their conclusions, as results of the current paper showed low correlation values between raw CAMS and measurements PM2.5 concentrations. However, after applying the LightGBM calibration technique, the correlations

increased significantly, and RMSE values were low, indicating that the calibrated PM2.5 estimates were more accurate.

Furthermore, the present results are compatible with the findings of (Jin *et al.*, 2022), who stressed the possibility of distinct calibration schemes to improve CAMS products. After calibration, the rise in coefficient of determination (R^2) values demonstrates the reduction in modelling biases and the improved performance of the calibrated PM2.5 data.

Overall, the calibration results using LightGBM algorithm are consistent with prior studies, suggesting that calibration approaches can significantly increase the accuracy and reliability of CAMS PM2.5 estimates. The calibrated PM2.5 data better matches with ground-based observations by correcting overestimation and lowering modelling biases, giving more reliable information for air quality assessments and decision-making processes. These findings emphasize the importance of calibration in increasing the utility and reliability of CAMS PM2.5 data for environmental monitoring and public health activities.

5 Conclusions and recommendations

The study of PM pollutants is a broad field with many branches, from studying the chemical composition of the PM particles to improving measurements and developing computer-based models to forecast PM pollution of any kind. The research conducted in this PhD thesis has contributed valuable insights into the impact of obstacles, hills, and Saharan dust storms on PM concentrations and the use of satellite-based models and machine learning algorithms to estimate PM concentrations and calibrate CAMS PM2.5 data. Based on the findings of this research, the following conclusions and recommendations can be made:

The study of the effects of simple obstacles on the PM10 concentration provides valuable insights into the effects of obstacle height, distance from the source, and wind speed on PM10 concentration in different sensor locations. The study's results indicate that obstacles can significantly impact PM10 concentration levels, with higher obstacle heights and greater distances from the source leading to decreased PM10 concentration levels. Additionally, the study found that wind speed plays a crucial role in PM10 concentration levels, with higher wind speeds leading to higher PM10 concentration levels. The study's findings are consistent with previous research on the effects of urban architecture and wind speed on PM concentrations. For example, a study on the influence of wind speed on airflow and fine particle transport within different building layouts of an industrial city found that height variation and layout of urban architecture affect the local concentration distribution of PM (Mei *et al.*, 2018). Similarly, another study on the effects of windbreaks on particle concentrations from agricultural fields under various wind conditions showed that vegetation barriers can alter particle transport by affecting airflow (Chang *et al.*, 2019). In conclusion, the results of the research paragraph provide valuable insights into the effects of obstacle height, distance from the source, and wind speed on PM10 concentration levels. By considering the impact of obstacles and wind speed on PM10 concentration levels, policymakers and urban planners can develop effective strategies to minimise the impact of PM10 on public health in urban environments.

The study the impact of hill elevation on the dispersion of PM plumes showed that the correlation between PM10 concentration and wind speed at different sensor locations was affected. The results show that hills can significantly affect the dispersion pathway of PM particles, with different slopes creating other flows of PM dispersion. The study found that the correlation between PM10 concentration and wind speed decreases as the height of the hill increases, indicating changes in the wind flow created by different elevations of the hill and that, in general, topography can significantly impact the level of PM concentrations. In a study done by Wen *et al.* (2022), they discuss the quantitative disentanglement of topography's geographical impacts on PM2.5 pollution in China. They emphasise that mountains significantly impact the spatial heterogeneity of PM2.5 pollution levels. The study found that high-altitude mountains and plateaus experience lower levels of PM2.5 pollution, while plains and surrounding platforms and hills suffer from severe pollution. Also, the mountain's blocking effects begin to play an efficient role when their altitudes reach a specific value; however, the exact altitude values vary by different mountains, with a value of 163 m for all typical mountains with absolute PM2.5 concentration differences between their two sides greater than 10 $\mu g/m^3$. Even though the experiments included relatively low height, it showed that height can alter the PM concentrations, even in short range. In conclusion, the research results provide valuable insights into the effects of hill elevation on the dispersion of PM plumes and the correlation between PM10 concentration and wind speed at different sensor locations.

The study of Saharan dust storms in Hungary revealed that they increased PM10 and PM2.5 concentration levels, and the seasonality and frequency are changing. Varga (2020) indicates numerous intense events happened after 2014 when an unusually significant amount of mineral dust was washed out. All occurred between the end of October and February, and the increase in

frequency and intensity of wintertime dust depositional events in the Carpathian Basin (Hungary mainly) is attributed to climatic conditions. Our study for the period between 2018 and 2022 showed changes in the frequency of the intense Saharan dust storms in Hungary (more likely to happen between February and April) and also the intensity (Recent March 2022 as an example), and dominant Type-2 events that are connected to Central Mediterranean cyclones which is responsible for dust transport.

Machine learning is a powerful tool that can be used to estimate PM concentrations from MERRA-2 data. In the study, we estimate PM concentrations using two approaches, mainly coupling MERRA-2 AOD and five PM components with meteorological data and MERRA-2 AOD and in-situ measurements of primary air pollutants (SO_2, NO_2, O_3) with meteorological data. The best results were given using the Extra-tree machine learning algorithm in the two approaches for the three stations chosen for Hungary (Budapest, Kazincbarcika, and Kecskemet) with an R^2 between 0.73 and 0.84. Our models performed well for the selected stations compared to other studies that used MERRA-2 data to estimate PM concentrations using machine learning. Dhandapani, Iqbal and Kumar (2023) Apply machine learning models on MERRA2 data to predict surface PM2.5 concentrations over India. Overall, the study evaluates the utility of Machine Learning (ML) models, focusing on XGBoost (XGB), Random Forest (RF), and LightGBM (LGBM) individual models, as well as a stacking technique. The authors compared the performance of these models and found that the stacking technique (R^2= 0.77) outperformed unique models (R^2=0.73), showing the best hourly prediction in the eastern (R^2 = 0.80) and northern regions (R^2 = 0.63). In another study by Sayeed *et al.* (2022), they evaluated the performance of the machine learning model in estimating PM2.5 concentration, which outperformed the MERRA-2 empirical estimation of PM2.5 and exhibited a small and uniform bias throughout the day and in all seasons and proved to be better at estimating PM2.5 than the MERRA-2 practical calculations. In conclusion, Coupling MERRA-2 and meteorological data with other meaningful parameters and employing machine learning to predict PM concentrations can yield promising results, as demonstrated in our case for the three stations in Hungary.

For the calibration of CAMS PM2.5 data using machine learning in Hungary, our model improved the degree of accuracy of CAMS PM2.5 data from low R (>0.25) to higher R (<0.79), demonstrating the effectiveness of calibration schemes using LightGBM machine learning model in reducing modelling biases and enhancing the performance of CAMS PM2.5 data in Hungary. Jin *et al.* (2022) proposes a calibration method to improve the accuracy of CAMS PM2.5 data, using the Extremely Tree machine learning model, resulting in significant accuracy improvement, with R reaching up to 0.81 and RMSE decreasing by about 60% for the original CAMS PM2.5 For China, US and Africa. Overall, CAMS reanalysis datasets require significant improvement for use in local and regional air quality monitoring, and our study showed a significantly better correlation between the Calibrated PM2.5 and in-situ measurements of PM2.5 over Hungary, suggesting an improved accuracy of Calibrated CAMS PM2.5.

Concerning the recommendations, further research on the impact of obstacles on PM concentrations: It is recommended that the impact of different types of obstacles on PM concentrations be understood and that more accurate models to predict PM concentrations in the presence of obstacles be developed.

Use of satellite-based models to estimate PM pollutants: It is recommended to use satellite-based models to estimate PM pollutants in Hungary and to compare the results with in situ measurements to validate the accuracy of the models.

Further research on the impact of Saharan dust storms on PM concentrations: It is recommended that the impact of different types of dust storms on PM concentrations be understood and that more accurate models to predict PM concentrations during dust storms be developed.

Use of machine learning algorithms to estimate PM concentrations: It is recommended to use machine learning algorithms to estimate PM concentrations in other locations in Hungary and to expand the study by using more extensive data.

In conclusion, the research conducted in this PhD thesis has contributed valuable insights into the impact of various factors on PM concentrations. It has provided recommendations for further research to improve the accuracy of PM concentration predictions. This research's findings can help protect public health and the environment by providing accurate and reliable PM concentration data.

6 New scientific results

1. *As a result of small-scale experiments, I have shown that the PM10 concentration from a point source depends on the wind speed (Ws=0-2.9 m/s) and the height and position of a simple "obstacle" placed between the source and the measurement point - with a critical ratio of obstacle distance from the source over obstacle height ranging (OD/OH) from 0 to 2.3. A linear equation established the relationship between the mentioned parameters to calculate PM10 concentration, reaching an R^2 of 0.79. I declare that, the obstacle's position emerges as an essential determinant shaping the estimated PM10 concentration, underscoring its profound significance in our findings.*

Research conducted a multiple regression analysis to predict $PM10_A$ concentration using "OH" (Obstacle height), "OD" (Distance of the obstacle from the source), "$PM10_C$" (PM10 concentration), and "Ws" (Wind speed) as independent variables. The regression model equation was:

$$PM10_A = 143.07 - 71.86 * OH - 171.42 * OD + 1.23 * PM10_C + 12.34 * Ws$$

Results showed a moderately significant positive connection between $PM10_A$ and the independent variables (R=0.89). The model explained 79% of $PM10_A$ variation (R^2=0.79). Obstacle height, distance from the source, and wind speed had significant positive effects on $PM10_A$.
The experiment elicits know how a simple obstacle in the form of a solid barrier can reduce the PM concentrations. And using the equation, it is possible to strategically position the obstacle to obstruct the PM particle movements and opt for the optimal possible height to effectively isolate the area from direct PM transportation to minimize the PM concentration levels to acceptable levels.

2. *I conclude that hill height influences PM10-wind speed correlations negatively as a result of a series of outdoor trials investigating the effects of different ground surface elevations (H=0-1 m) on the dispersion of PM10 and PM2.5 pollutants at various wind speeds (Ws=0-6 m/s) in short range, revealing the complicated interplay between topography and air pollution patterns.*

   The study examines the correlations between PM10 concentrations recorded by three different sensors (S1, S2, and S3) over three different experimental situations with varied wind speeds on a hill with varying heights. Correlations between PM10 concentrations and wind speed differ between sensors (S1, S2, S3) and experimental cases on a hill with variable elevation.
   Sensor 1 (S1), which is near to the source, has a negative correlation of -0.5 in Case 3, but weaker and inverted correlations to -0.18 and 0.2 in Cases 2 and 1 respectively. Sensor 2 (S2) positioned at the edge of the hill's slope, has continuous positive correlations in all cases, however they are weaker in case 2 (0.27) and 3 (0.18), probably due to changes in nearby terrain. Sensor 3 (S3), located atop the hill, retains positive associations, although they decline as elevation increases, yielding values of 0.65 (Case 3), 0.84 (Case 2), and 0.8 (Case 1), suggesting that shifting wind patterns impact PM10 transport.

3. *HYSPLIT dust simulations offer compelling insights into the origin and trajectory of Saharan dust particles. Our analysis reveals that these particles observed in the Gulf of Mexico during the June 2020 Saharan storm unmistakably trace back to the Moroccan and Mauritanian Saharan regions. Dust storms emerged from specific hotspots, such as Tinduf near the Moroccan border and Adrar, Tiris Zemmour, and Tagant in Mauritania. Markedly, the June 2020 Saharan dust storm was associated with the highest June aerosol optical depths recorded, exceeding AOD=3.5 in Bir Anzarane, Morocco, and an astonishing AOD=5.5 in Nouakchott, Mauritania, affecting the PM concentrations to unhealthy levels in several US Gulf States, further substantiating that due to climate change, the Saharan dust storms are getting more intense, especially in the Moroccan and Mauritanian Sahara.*

   Based on HYSPLIT dust simulation and also comparing the simulation results maps with MODIS AOD average maps, the dust storm in the region of Morocco and Mauritania started on the 14th of June 2020, from the region of Tinduf close to the borders of Morocco with Algeria, Adrar Tiris Zemmour and Tagant in Mauritania. Due to the wind field heading towards the Atlantic oceans, the dust was transported across the ocean to the American continents. In addition, HYSPLIT cluster analysis from many places in Morocco and Mauritania showed a significant percentage of PM10 particles that negatively affected the PM10 and PM2.5 concentrations in the Caribbean Sea and US coastal in the Gulf of Mexico, originated from places like Bir Anzarane, Morocco, Nouakchott, and Tichit Mauritania, and Bordj Badji Mokhtar Algeria. In addition, there was an increase in Bir Anzarane, Morocco, in AOD with a value of 3.522. This marks a surge of 188% compared to the highest recorded value between 2010 and 2019, 1.87 in June 2017. Similarly, in Nouakchott, Mauritania, the peak AOD value from 2010 to 2019 was 2.78 in June 2010. However, there was a rise in June 2020, with an AOD value reaching 5.87, representing an increase of around 211%.

4. *I conclude that events where the Saharan dust transport was brought on by Central Mediterranean cyclones to Hungary, - called type 2 Saharan dust storm events - were dominant in 2018 and 2022 and usually happen in February, March, or April, with a maximum hourly dust mass between 450 and 1000 mg/m2. The 2 Saharan Dust storms in March 2022 raised the concentrations of PM10 and PM2.5 in Budapest by 12 µg/m3 and 10 µg/m3, respectively, during the first Saharan Dust event and by 14 µg/m3 and 5 µg/m3 during the Second Saharan Dust event.*

   Based on my evaluation of the Saharan dust storm events in Hungary, between 2018 and 2022, 11 SDEs were identified in Hungary. Type 2 SDEs were dominated in that period and characterised by high Dust mass, negatively affecting the PM concentrations. And most of the time, the SDEs were likely to occur between February and April. Moreover, March 2022 was a unique month due to two extreme outbreaks of Saharan Dust events (14-19 and 28-31), which were unusual throughout the study period. In general, the Saharan dust events between 2018 and 2022 were associated with an increase of PM10 daily average concentration by a factor of 2 or more, according to PM10 concentration measurements from an urban background air quality station in Budapest.

5. *I find that PM10 and PM2.5 concentrations simulated from MERRAero data, encompassing five PM species (SO_4, OC, BC, DS and SS), AOD, and meteorological parameters (T, Ws_{10}, Ws_{50}, RH, P and PBLH), between 2019 and 2021 accurately estimated using the Extra-Tree regression model for three cities in Hungary (Budapest, Kecskemét and Kazincbarcika), achieving R^2 values between 0.75 and 0.8 for PM10 and PM2.5.*

Based on my evaluation of the estimated PM10 and PM2.5 based on the five PM species simulated by the MERRAero hourly data collection of (SO_4, OC, BC, DS, and SS), in conjunction with aerosol optical depth (AOD) and meteorological parameters (T, Ws_{10}, Ws_{50}, RH, P, and PBLH) for the period spanning 2019 to 2021, using a variety of machine learning algorithms, it is discerned that the Extra-Tree regression model consistently produced the most favourable outcomes. The quantitative results, disaggregated by location, are as follows:

- In Budapest, the determination coefficient (R^2) reached 0.78 and 0.8 for the estimation of PM10 and PM2.5, respectively.
- In Kecskemét, the R^2 values achieved were 0.75 and 0.76 for PM10 and PM2.5 estimation, respectively.
- For Kazincbarcika, the R^2 values obtained for PM10 and PM2.5 were 0.77 and 0.8, respectively.

The significance of these results lies in their potential to enhance air quality monitoring and forecasting in urban areas such as Budapest, Kecskemét, and Kazincbarcika. The Extra-Tree regression model demonstrates robust predictive capabilities, with R^2 values consistently approaching or exceeding 0.75.

6. *Utilizing Machine learning algorithm (Extra-Tree regression model) to estimate PM2.5 concentrations based on MERRA-2 AOD, meteorological data (T, Ws_{10}, Ws_{50}, RH, P and PBLH), and in-situ measurements of NO_2, O_3, and SO_2 over three years (2019 to 2021) in 3 locations in Hungary (Budapest, Kecskemét and Kazincbarcika) underscores the importance of machine learning in PM2.5 prediction attaining an R^2 ranging from 0.73 to 0.83, and RMSE between 5.5 and 7.6 µg/m^3.*

   The utilisation of machine learning algorithms to estimate PM2.5 concentrations based on a comprehensive dataset comprising MERRA-2 AOD, meteorological data, and in-situ measurements of NO2, O3, and SO2 over three years (2019 to 2021) for 3 locations in Hungary underscored the effectiveness of machine learning as a valuable predictive tool but also revealed the superiority of the Extra-Tree regression model over alternative approaches. Specific results for each location are as follows:

   - In case of Budapest, I had an R^2 of 0.73, RMSE of 5.5 µg/m^3, and MAE of 3.7 µg/m^3
   - In case of Kecskemét, I had an R^2 of 0.75, RMSE of 6.4 µg/m^3, and MAE of 4.2 µg/m^3.
   - In case of Kazincbarcika, I had an R^2 of 0.84, RMSE of 7.6 µg/m^3, and MAE of 4.8 µg/m^3.

   These findings hold paramount importance as they affirm the applicability of machine learning for precise PM2.5 predictions, offering a robust and versatile methodology for air quality assessment and prediction in these specific geographical areas. Such accurate

predictive models are instrumental for public health, urban planning, and environmental management

7. *Using the LightGBM algorithm in calibrating CAMS PM2.5 data for 11 air quality stations in Hungary reveals a remarkable improvement in data accuracy and alignment with in-situ measurements with post-calibration, correlations substantially increased, with values ranging from 0.78 to 0.88, underscoring a solid association between calibrated CAMS data and actual PM2.5 measurements, and a coefficient of determination values ranging from 0.61 to 0.77.*

   - Correlation analysis shows initial alignment between raw CAMS data and in-situ measurements, with correlations before training ranging from 0.071 to 0.207. After training, correlations significantly improve, ranging from 0.787 to 0.882, demonstrating a strong association between calibrated CAMS data and in-situ PM2.5 measurements.
   - The coefficient of determination values ranges from 0.618 to 0.774, indicating a substantial portion of the variance in in-situ PM2.5 measurements is explained by the calibrated CAMS PM2.5.
   - Lower root mean square error values reflect reduced discrepancies between the calibrated CAMS PM2.5 and actual measurements, indicating improved accuracy and precision.

   The findings underscore the critical role of calibration in improving the accuracy of air quality data (such as CAMS PM raw data). Enhanced correlations, higher coefficient of determination values, and reduced root mean square error values following machine learning calibration are scientifically significant and have direct practical implications.

7 SUMMARY

Studying the PM pollutants is broad field that have many branches, from studying the chemical composition of the PM particles to improving measurements and developing computer-based models to forecast PM pollutions of any kind. In the initial phase of my research, I looked over the literature in a few relevant subfields, which led to Saharan dust storm study both the June 2020 event and to know the effect of climate change on the triggering and transport of Saharan dust to Hungary. In addition to understand the dispersion of PM particles around simple obstacle and in small range elevated hills, as well as understanding the relationship between inside and outside PM concentrations. Moreover, deep search in the literature lights the fact that no one has done an estimation of the PM pollutants in Hungary using the Satellite based models, despite the fact of the increasing interest in this subfield of research and the rise of the number of papers published in order to improve the use of the Satellite datasets to estimate one of the major and dangerous air pollutants.

The study of small-scale PM dispersion around simple obstacle demonstrated that Obstacle height, Distance of the obstacle from the source, and Wind speed had significant positive effects on PM10 concentration after the obstacle. The analysis revealed a moderately significant positive connection between the dependent variable (PM10 concentration after the obstacle) and the set of independent factors, as indicated by the correlation coefficient (R) of 0.89. Moreover, the independent variables in the model collectively explained approximately 79% of the variation in the dependent variable, as reflected by the coefficient of determination (R^2) of 0.79. Overall, the research provides valuable insights into the impact of obstacle height, distance from the source, and wind speed on PM10 concentration and confirms the transport behaviour of PM particles in both small-scale experiments and larger-scale urban settings.

For the study of the effects of small hills on PM concentrations, the results revealed that at low wind speeds (0 and 0.7 m/s), the average concentrations of PM10 and PM2.5 were similar for all three cases. However, at higher wind speeds (2.4, 3.7, and 5.1 m/s), the average concentrations of PM10 and PM2.5 were significantly higher in the 1m height and 0.8 m height cases compared to the flat ground surface. Furthermore, the study showed that the difference in ground surface elevation between the 1m height and 0.8 m height cases had a notable impact on PM dispersion. The elevated ground surface (hill) altered the dispersion pathways of PM particles, resulting in higher concentrations in certain areas. Sensor 1 recorded higher PM concentrations in cases 2 and 3 compared to case 1, especially at wind speeds below 3 m/s, primarily due to the reflective effect of the hill and low wind speeds. Sensor 2 registered higher PM concentrations before the hill, indicating particle trapping in that area. The study also employed multiple linear regression to estimate PM10 concentration at the top of the hill based on measurements from sensor near the source, and sensor at the bottom of the hill, wind speed, and hill height. The regression analysis showed a strong positive correlation (R=0.9) between the dependent variable (PM10 concentration at the top of the hill) and the combination of independent variables (mentioned above). Approximately 82% of the variance in PM10 concentration at the top of the hill was explained by the independent variables (R^2=0.82). Also, the correlation coefficient between measured PM10 by all three sensors and wind speed demonstrates that hill height is important in shaping correlations between PM10 and wind speed, revealing intricate connections between topography and air pollution pattern.

For the dust storm simulation over the Sahara Desert (Moroccan and Mauritanian regions) using HYSPLIT, the average PM10 concentration between 0 and 100m reached severe levels according to the HYSPLIT dust simulation results. Regions like Dakhla-Oued Ed-Dahab in Morocco, Adrar and Tiris Zemmour in Mauritania had higher PM10 concentrations (higher than 100 µg/m^3) and AOD values (between 0.7 and 1) during the 4 days of the dust storm. Moreover, PM10 particles

were transported over the Atlantic Ocean to the Caribbean Sea and the Gulf of Mexico, causing raise in the level of concentrations in those regions. The tropospheric level of the Caribbean Sea and the Gulf of Mexico was loaded by dust particles transported from the study area. Bir Anzarane Morocco, Nouakchott and Tichit Mauritania, and Bordj Badji Mokhtar Algeria all contributed to the high PM10 concentrations observed in the Martinique islands and the southern United States, while the top altitude of the dust layer was between 4 and 4.5 km, according to the backscatter vertical profile measured by CALIPSO. Therefore, PM10 concentration and AOD revealed their peak values during June 2020 dust storm, and this is evidenced by AOD values recorded at Bir Anzarane, Morocco, and Nouakchott, Mauritania, both of which are historical by June norms. There was a rise in AOD, with a value of 3.52 in Bir Anzarane, Morocco. This represents an increase of 188% above the highest recorded figure between 2010 and 2019, which was 1.87 in June 2017. Similarly, from 2010 to 2019, the AOD value in Nouakchott, Mauritania, was 2.78 in June 2010. However, there was a spike in June 2020, with an AOD value reaching as high as 5.87, reflecting a 211% increase.

For the identification and evaluation of the Saharan dust storm events in Budapest, Hungary between 2018 and 2022, type 2 SDEs predominated in that period with 5 occurrences, whereas types 1 and 3 appeared three times each. In addition, February, March, and April have seen the most SDEs (7 times), and SDEs occurring in those months are more likely to be severe events because that period had the highest dust mass recorded (SDE4 - April 23-27, 2019) and an increase in PM10 daily average concentration of at least a factor of 2.

For the case study of the Saharan dust effects on PM10 and PM2.5 concentrations in Budapest in March 2022, the two type 3 dust storms contributed to an increase in PM10 and PM2.5 concentration levels. The PM10 concentrations increased by 12 and 14 $\mu g/m^3$, during first and second Saharan dust events, while for PM2.5 the concentration rise by 10 and 5 $\mu g/m^3$, in first and second Saharan dust events respectively, highlighting the fact that first Saharan dust event had bigger impact on PM2.5 that the second Saharan dust events, in contrast for PM10, the impact of the two Saharan dust events were nearly similar.

For the evaluation of PM surface concentrations simulated by Version 5.12.4 of NASA's MERRA-2 Aerosol Reanalysis over Hungary in the period between 2019 and 2021, the estimation of the PM10 and PM2.5 concentrations done in two approaches. The first approach involved estimating PM10 and PM2.5 using equations 1 and 2, that calculate the PM10 and PM2.5 in function of BC, OC, DS, SO_4 and SS concentrations given by MERRA-2 Aerosol analysis dataset and compare it with real measurements of PM10 and PM2.5, in addition to estimations using machine learning algorithms such as MLR, OLS, RF, and XT, and the data used in machine learning algorithm is coupled with meteorological data (T, P, RH, Ws_{10}, Ws_{50} and PBLH) and AOD. The second approach used the machine learning techniques used in the first approach to estimate PM2.5 and this time based on AOD in conjunction with observations of NO_2, O_3, SO_2, and meteorological data (T, P, RH, Ws_{10}, Ws_{50} and PBLH). And, both first and second approaches were applicated in 3 cities in Hungary, Budapest, Kecskemét, and Kazincbarcika. In case of the first approaches, results showed that using XT model gave the best results for all the three locations of the study, for Budapest I got an R^2 of 0.78 and 0.8 for PM10 and PM2.5 estimations respectively, and for Kecskemét an R^2 of 0.75 and 0.76, in addition for an R^2 of 0.77 and 0.8 for PM10 and PM2.5 estimations respectively for Kazincbarcika, proving the effectiveness of the XT machine learning model in estimating the PM concentrations. Moreover, for the second approach, Estimating the PM2.5 using XT model also gave the best results. The best R^2 achieved was for Kazincbarcika with value of 0.84, followed by Kecskemét with value of 0.75 and Budapest with value of 0.73.

The use of Satellite based data, coupled with meteorological data can give accurate estimations of PM concentrations, especially PM2.5, where it is highlighted in many research studies that PM2.5

have a complex relationship with AOD and can be used to predict it's concentrations. Furthermore, the use of machine learning or deep learning methods prove to be useful tool in PM estimations in the study that I've done in this thesis, and my study can be expended by using bigger data and to other locations that could cover all the Hungarian territory.

Finally for the calibration of CAMS PM2.5 data, results reveal significant improvements in various metrics. The correlation coefficients before (ranged from 0.07 to 0.20) and after (ranging from 0.78 to 0.88) of the calibration method demonstrate noteworthy enhancements, indicating a stronger alignment between the CAMS PM2.5 data and in situ measurements. Additionally, the coefficient of determination (R^2) (ranged from 0.61 to 0.77) exhibits substantial increases, highlighting the improved predictive power of the calibrated data. The calibration process also leads to reductions root mean squared error (RMSE), indicating decreased variability between predicted and observed PM2.5 values.

These calibration outcomes have implications not only for Hungary but also for other countries grappling with air quality issues. Accurate and reliable CAMS PM2.5 data serves as a vital resource for governments, environmental agencies, and health organizations worldwide. By leveraging calibration techniques like LightGBM, countries can enhance the quality of their air quality datasets, leading to more accurate assessments of pollution levels and better-informed decision-making.

8 Appendices

8.1 Bibliography

Achilleos, S. *et al.* (2014) 'PM10 concentration levels at an urban and background site in Cyprus: The impact of urban sources and dust storms', *Journal of the Air & Waste Management Association*, 64(12), pp. 1352–1360. Available at: https://doi.org/10.1080/10962247.2014.923061.

Alfano, B. *et al.* (2020) 'A Review of Low-Cost Particulate Matter Sensors from the Developers' Perspectives', *Sensors (Basel, Switzerland)*, 20(23), p. 6819. Available at: https://doi.org/10.3390/s20236819.

Ali, Md.A. *et al.* (2022) 'Accuracy assessment of CAMS and MERRA-2 reanalysis PM2.5 and PM10 concentrations over China', *Atmospheric Environment*, 288, p. 119297. Available at: https://doi.org/10.1016/j.atmosenv.2022.119297.

Ashrafi, K. *et al.* (2014) 'Dust storm simulation over Iran using HYSPLIT', *Journal of Environmental Health Science and Engineering*, 12(1), p. 9. Available at: https://doi.org/10.1186/2052-336X-12-9.

Austin, E. *et al.* (2015) 'Laboratory Evaluation of the Shinyei PPD42NS Low-Cost Particulate Matter Sensor', *PLOS ONE*, 10(9), p. e0137789. Available at: https://doi.org/10.1371/journal.pone.0137789.

Baker, J. (2010) 'A cluster analysis of long range air transport pathways and associated pollutant concentrations within the UK', *Atmospheric Environment*, 44(4), pp. 563–571. Available at: https://doi.org/10.1016/j.atmosenv.2009.10.030.

Baron, P.A., Kulkarni, P. and Willeke, K. (eds) (2011) *Aerosol measurement: principles, techniques, and applications*. 3rd ed. Hoboken, N.J: Wiley.

Báthory, C. *et al.* (2019) 'Stationary and mobile particulate matter concentration measurement in Miskolc', *International Journal of Engineering and Management Sciences*, 4(4), pp. 338–344. Available at: https://doi.org/10.21791/IJEMS.2019.4.38.

Báthory, C. *et al.* (2022) 'Low-cost monitoring of atmospheric PM—development and testing', *Journal of Environmental Management*, 304, p. 114158. Available at: https://doi.org/10.1016/j.jenvman.2021.114158.

Belle, J. and Liu, Y. (2016) 'Evaluation of Aqua MODIS Collection 6 AOD Parameters for Air Quality Research over the Continental United States', *Remote Sensing*, 8(10), p. 815. Available at: https://doi.org/10.3390/rs8100815.

Bozlaker, A. *et al.* (2013) 'Quantifying the Contribution of Long-Range Saharan Dust Transport on Particulate Matter Concentrations in Houston, Texas, Using Detailed Elemental Analysis', *Environmental Science & Technology*, p. 130909083424001. Available at: https://doi.org/10.1021/es4015663.

Bozlaker, A. *et al.* (2019) 'Identifying and Quantifying the Impacts of Advected North African Dust on the Concentration and Composition of Airborne Fine Particulate Matter in Houston and

Galveston, Texas', *Journal of Geophysical Research: Atmospheres*, 124(22), pp. 12282–12300. Available at: https://doi.org/10.1029/2019JD030792.

Breiman, L. (2001) 'Random Forests', *Machine Learning*, 45(1), pp. 5–32. Available at: https://doi.org/10.1023/A:1010933404324.

Brook, R.D. *et al.* (2010) 'Particulate Matter Air Pollution and Cardiovascular Disease', *Circulation*, 121(21), pp. 2331–2378. Available at: https://doi.org/10.1161/CIR.0b013e3181dbece1.

Bruckmann, P. *et al.* (2008) 'An outbreak of Saharan dust causing high PM(10) levels north of the Alps', *GEFAHRSTOFFE REINHALTUNG DER LUFT*, 68, pp. 490–498.

Brunekreef, B. and Maynard, R.L. (2008) 'A note on the 2008 EU standards for particulate matter', *Atmospheric Environment*, 42(26), pp. 6425–6430. Available at: https://doi.org/10.1016/j.atmosenv.2008.04.036.

Buccolieri, R. *et al.* (2022) 'Obstacles influence on existing urban canyon ventilation and air pollutant concentration: A review of potential measures', *Building and Environment*, 214, p. 108905. Available at: https://doi.org/10.1016/j.buildenv.2022.108905.

Buchard, V. *et al.* (2014) 'Evaluation of GEOS-5 sulfur dioxide simulations during the Frostburg, MD 2010 field campaign', *Atmospheric Chemistry and Physics*, 14(4), pp. 1929–1941. Available at: https://doi.org/10.5194/acp-14-1929-2014.

Buchard, V. *et al.* (2016) 'Evaluation of the surface PM2.5 in Version 1 of the NASA MERRA Aerosol Reanalysis over the United States', *Atmospheric Environment*, 125, pp. 100–111. Available at: https://doi.org/10.1016/j.atmosenv.2015.11.004.

Buchard, V. *et al.* (2017) 'The MERRA-2 Aerosol Reanalysis, 1980 Onward. Part II: Evaluation and Case Studies', *Journal of Climate*, 30(17), pp. 6851–6872. Available at: https://doi.org/10.1175/JCLI-D-16-0613.1.

Butt, E.W. *et al.* (2016) 'The impact of residential combustion emissions on atmospheric aerosol, human health, and climate', *Atmospheric Chemistry and Physics*, 16(2), pp. 873–905. Available at: https://doi.org/10.5194/acp-16-873-2016.

Cao, H. *et al.* (2015) 'Identification of dust storm source areas in West Asia using multiple environmental datasets', *Science of The Total Environment*, 502, pp. 224–235. Available at: https://doi.org/10.1016/j.scitotenv.2014.09.025.

Chaboureau, J.-P. *et al.* (2016) 'Fennec dust forecast intercomparison over the Sahara in June 2011', *Atmospheric Chemistry and Physics*, 16(11), pp. 6977–6995. Available at: https://doi.org/10.5194/acp-16-6977-2016.

Chang, X. *et al.* (2019) 'Effect of windbreaks on particle concentrations from agricultural fields under a variety of wind conditions in the farming-pastoral ecotone of northern China', *Agriculture, Ecosystems & Environment*, 281, pp. 16–24. Available at: https://doi.org/10.1016/j.agee.2019.04.017.

Chapman, R. and Gasparovic, R. (2022) *Remote Sensing Physics: An Introduction to Observing Earth from Space*. John Wiley & Sons.

Chen, G. *et al.* (2018) 'A machine learning method to estimate PM2.5 concentrations across China with remote sensing, meteorological and land use information', *Science of The Total Environment*, 636, pp. 52–60. Available at: https://doi.org/10.1016/j.scitotenv.2018.04.251.

Chen, H. *et al.* (2022) 'Effects of air pollution on human health – Mechanistic evidence suggested by in vitro and in vivo modelling', *Environmental Research*, 212, p. 113378. Available at: https://doi.org/10.1016/j.envres.2022.113378.

Cheng, Z. *et al.* (2016) 'Status and characteristics of ambient PM2.5 pollution in global megacities', *Environment International*, 89–90, pp. 212–221. Available at: https://doi.org/10.1016/j.envint.2016.02.003.

Cohen (2013) *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences*. 3rd edn. Routledge. Available at: https://doi.org/10.4324/9780203774441.

Copernicus (2020) *CAMS global reanalysis (EAC4)*. Available at: https://ads.atmosphere.copernicus.eu/cdsapp#!/dataset/cams-global-reanalysis-eac4?tab=overview.

Correia, A.W. *et al.* (2013) 'The Effect of Air Pollution Control on Life Expectancy in the United States: An Analysis of 545 US counties for the period 2000 to 2007', *Epidemiology (Cambridge, Mass.)*, 24(1), pp. 23–31. Available at: https://doi.org/10.1097/EDE.0b013e3182770237.

Cruz, J.A. *et al.* (2021) 'Strong links between Saharan dust fluxes, monsoon strength, and North Atlantic climate during the last 5000 years', *Science Advances*, 7(26), p. eabe6102. Available at: https://doi.org/10.1126/sciadv.abe6102.

Cuesta, J. *et al.* (2020) 'Three-dimensional pathways of dust over the Sahara during summer 2011 as revealed by new Infrared Atmospheric Sounding Interferometer observations', *Quarterly Journal of the Royal Meteorological Society*, 146(731), pp. 2731–2755. Available at: https://doi.org/10.1002/qj.3814.

Darwin, C. (1846) 'An account of the Fine Dust which often falls on Vessels in the Atlantic Ocean.', *Quarterly Journal of the Geological Society*, 2(1–2), pp. 26–30. Available at: https://doi.org/10.1144/GSL.JGS.1846.002.01-02.09.

Delany, A.C. *et al.* (1967) 'Airborne dust collected at Barbados', *Geochimica et Cosmochimica Acta*, 31(5), pp. 885–909. Available at: https://doi.org/10.1016/S0016-7037(67)80037-1.

Dhandapani, A., Iqbal, J. and Kumar, R.N. (2023) 'Application of machine learning (individual vs stacking) models on MERRA-2 data to predict surface PM2.5 concentrations over India', *Chemosphere*, 340, p. 139966. Available at: https://doi.org/10.1016/j.chemosphere.2023.139966.

Diette, G.B. *et al.* (2012) 'OBSTRUCTIVE LUNG DISEASE AND EXPOSURE TO BURNING BIOMASS FUEL IN THE INDOOR ENVIRONMENT', *Global heart*, 7(3), pp. 265–270. Available at: https://doi.org/10.1016/j.gheart.2012.06.016.

Dobrea, M. *et al.* (2020) 'Machine Learning algorithms for air pollutants forecasting', in *2020 IEEE 26th International Symposium for Design and Technology in Electronic Packaging (SIITME). 2020 IEEE 26th International Symposium for Design and Technology in Electronic Packaging (SIITME)*, Pitesti, Romania: IEEE, pp. 109–113. Available at: https://doi.org/10.1109/SIITME50350.2020.9292238.

Does, M. *et al.* (2016) 'Particle size traces modern Saharan dust transport and deposition across the equatorial North Atlantic', *Atmospheric Chemistry and Physics*, 16, pp. 13697–13710. Available at: https://doi.org/10.5194/acp-16-13697-2016.

Draxler, R., Ginoux, P. and Stein, A. (2010) 'An empirically derived emission algorithm for wind-blown dust', *Journal of Geophysical Research*, 115. Available at: https://doi.org/10.1029/2009JD013167.

Draxler, R. and Hess, G. (1998) 'An overview of the HYSPLIT_4 modeling system for trajectories, dispersion, and deposition', *Australian Meteorological Magazine*, 47, pp. 295–308.

Draxler, R.R. *et al.* (2001) 'Estimating PM10 air concentrations from dust storms in Iraq, Kuwait and Saudi Arabia', *Atmospheric Environment*, 35(25), pp. 4315–4330. Available at: https://doi.org/10.1016/S1352-2310(01)00159-5.

Escudero, M. *et al.* (2006) 'Determination of the contribution of northern Africa dust source areas to PM10 concentrations over the central Iberian Peninsula using the Hybrid SingleParticle Lagrangian Integrated Trajectory model (HYSPLIT) model', *Journal of Geophysical Research*, 111. Available at: https://doi.org/10.1029/2005JD006395.

European Commission. Joint Research Centre. (2018) *World atlas of desertification: rethinking land degradation and sustainable land management*. LU: Publications Office. Available at: https://data.europa.eu/doi/10.2760/06292 (Accessed: 28 January 2023).

European Environment Agency (2022) *Europe's air quality status 2022*. LU: Publications Office (EEA Briefing). Available at: https://data.europa.eu/doi/10.2800/049755 (Accessed: 31 January 2023).

European Environment Agency. (2022) *Hungary – air pollution country fact sheet — European Environment Agency*. Available at: https://www.eea.europa.eu/themes/air/country-fact-sheets/2022-country-fact-sheets/hungary-air-pollution-country.

Evan, A.T. *et al.* (2016) 'The past, present and future of African dust', *Nature*, 531(7595), pp. 493–495. Available at: https://doi.org/10.1038/nature17149.

Fang, Y. *et al.* (2013) 'Air pollution and associated human mortality: the role of air pollutant emissions, climate change and methane concentration increases from the preindustrial period to present', *Atmospheric Chemistry and Physics*, 13(3), pp. 1377–1394. Available at: https://doi.org/10.5194/acp-13-1377-2013.

Ferenczi, Z. and Bozó, L. (2017) 'Effect of the long-range transport on the air quality of greater Budapest area', *International Journal of Environment and Pollution*, 62(2/3/4), p. 407. Available at: https://doi.org/10.1504/IJEP.2017.089428.

Ferenczi, Z., Homolya, E. and Bozó, L. (2020) 'Detailed Assessment of a Smog Situation Detected in the Sajó Valley, Hungary', in C. Mensink, W. Gong, and A. Hakami (eds) *Air Pollution Modeling and its Application XXVI*. Cham: Springer International Publishing (Springer Proceedings in Complexity), pp. 351–356. Available at: https://doi.org/10.1007/978-3-030-22055-6_56.

Fink, A.H. and Reiner, A. (2003) 'Spatiotemporal variability of the relation between African Easterly Waves and West African Squall Lines in 1998 and 1999', *Journal of Geophysical Research: Atmospheres*, 108(D11). Available at: https://doi.org/10.1029/2002JD002816.

Fluck, E. and Raveh-Rubin, S. (2023) 'Dry air intrusions link Rossby wave breaking to large-scale dust storms in Northwest Africa: Four extreme cases', *Atmospheric Research*, 286, p. 106663. Available at: https://doi.org/10.1016/j.atmosres.2023.106663.

Fong, K.C. *et al.* (2019) 'Fine Particulate Air Pollution and Birthweight: Differences in Associations along the Birthweight Distribution', *Epidemiology (Cambridge, Mass.)*, 30(5), pp. 617–623. Available at: https://doi.org/10.1097/EDE.0000000000001039.

Francis, D. *et al.* (2020) 'The Atmospheric Drivers of the Major Saharan Dust Storm in June 2020', *Geophysical Research Letters*, 47(24). Available at: https://doi.org/10.1029/2020GL090102.

Francis, D. *et al.* (2022) 'Atmospheric rivers drive exceptional Saharan dust transport towards Europe', *Atmospheric Research*, 266, p. 105959. Available at: https://doi.org/10.1016/j.atmosres.2021.105959.

Fu, H. and Chen, J. (2017) 'Formation, features and controlling strategies of severe haze-fog pollutions in China', *Science of The Total Environment*, 578, pp. 121–138. Available at: https://doi.org/10.1016/j.scitotenv.2016.10.201.

Ganor, E. and Mamane, Y. (1982) 'Transport of Saharan dust across the eastern Mediterranean', *Atmospheric Environment (1967)*, 16(3), pp. 581–587. Available at: https://doi.org/10.1016/0004-6981(82)90167-6.

Gao, M., Cao, J. and Seto, E. (2015) 'A distributed network of low-cost continuous reading sensors to measure spatiotemporal variations of PM2.5 in Xi'an, China', *Environmental Pollution*, 199, pp. 56–65. Available at: https://doi.org/10.1016/j.envpol.2015.01.013.

Géron, A. (2022) *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow*. O'Reilly Media, Inc.

Getzewich, B.J. *et al.* (2018) 'CALIPSO lidar calibration at 532 nm: version 4 daytime algorithm', *Atmospheric Measurement Techniques*, 11(11), pp. 6309–6326. Available at: https://doi.org/10.5194/amt-11-6309-2018.

Geurts, P., Ernst, D. and Wehenkel, L. (2006) 'Extremely randomized trees', *Machine Learning*, 63(1), pp. 3–42. Available at: https://doi.org/10.1007/s10994-006-6226-1.

Ghahremanloo, M. *et al.* (2021) 'Estimating Daily High-Resolution PM2.5 Concentrations over Texas: Machine Learning Approach', *Atmospheric Environment*, 247, p. 118209. Available at: https://doi.org/10.1016/j.atmosenv.2021.118209.

Ginoux, P. *et al.* (2001) 'Sources and distributions of dust aerosols simulated with the GOCART model', *Journal of Geophysical Research: Atmospheres*, 106(D17), pp. 20255–20273. Available at: https://doi.org/10.1029/2000JD000053.

Giordano, M.R. *et al.* (2021) 'From low-cost sensors to high-quality data: A summary of challenges and best practices for effectively calibrating low-cost particulate matter mass sensors',

Journal of Aerosol Science, 158, p. 105833. Available at: https://doi.org/10.1016/j.jaerosci.2021.105833.

Gomes, J., Esteves, H. and Rente, L. (2022) 'Influence of an Extreme Saharan Dust Event on the Air Quality of the West Region of Portugal', *Gases*, 2(3), pp. 74–84. Available at: https://doi.org/10.3390/gases2030005.

Goudie, A.S. and Middleton, N.J. (2001) 'Saharan dust storms: nature and consequences', *Earth-Science Reviews*, 56(1), pp. 179–204. Available at: https://doi.org/10.1016/S0012-8252(01)00067-8.

Guaita, R. *et al.* (2011) 'Short-term impact of particulate matter (PM(2.5)) on respiratory mortality in Madrid', *International Journal of Environmental Health Research*, 21(4), pp. 260–274. Available at: https://doi.org/10.1080/09603123.2010.544033.

Guerreiro, C.B.B., Foltescu, V. and de Leeuw, F. (2014) 'Air quality status and trends in Europe', *Atmospheric Environment*, 98, pp. 376–384. Available at: https://doi.org/10.1016/j.atmosenv.2014.09.017.

Gueymard, C.A. and Yang, D. (2020) 'Worldwide validation of CAMS and MERRA-2 reanalysis aerosol optical depth products using 15 years of AERONET observations', *Atmospheric Environment*, 225, p. 117216. Available at: https://doi.org/10.1016/j.atmosenv.2019.117216.

Gupta, P. and Christopher, S.A. (2009a) 'Particulate matter air quality assessment using integrated surface, satellite, and meteorological products: 2. A neural network approach', *Journal of Geophysical Research: Atmospheres*, 114(D20). Available at: https://doi.org/10.1029/2008JD011497.

Gupta, P. and Christopher, S.A. (2009b) 'Particulate matter air quality assessment using integrated surface, satellite, and meteorological products: Multiple regression approach', *Journal of Geophysical Research: Atmospheres*, 114(D14). Available at: https://doi.org/10.1029/2008JD011496.

Harrison, R.M., Hester, R.E. and Querol, X. (eds) (2016) *Airborne Particulate Matter: Sources, Atmospheric Processes and Health*. Cambridge: Royal Society of Chemistry (Issues in Environmental Science and Technology). Available at: https://doi.org/10.1039/9781782626589.

He, L. *et al.* (2019) 'Assessment of MERRA-2 Surface PM2.5 over the Yangtze River Basin: Ground-based Verification, Spatiotemporal Distribution and Meteorological Dependence', *Remote Sensing*, 11(4), p. 460. Available at: https://doi.org/10.3390/rs11040460.

Hersbach, H. *et al.* (2020) 'The ERA5 global reanalysis', *Quarterly Journal of the Royal Meteorological Society*, 146(730), pp. 1999–2049. Available at: https://doi.org/10.1002/qj.3803.

Holstius, D.M. *et al.* (2014) 'Field calibrations of a low-cost aerosol sensor at a regulatory monitoring site in California', *Atmospheric Measurement Techniques*, 7(4), pp. 1121–1131. Available at: https://doi.org/10.5194/amt-7-1121-2014.

Hoyk, E., Kanalas, I. and Farkas, J.Z. (2020) 'Environmental changes in Kecskemét between 2009-2018', *Gradus*, 7(2), pp. 1–5. Available at: https://doi.org/10.47833/2020.2.AGR.001.

Hu, X. *et al.* (2017) 'Estimating PM2.5 Concentrations in the Conterminous United States Using the Random Forest Approach', *Environmental Science & Technology*, 51(12), pp. 6936–6944. Available at: https://doi.org/10.1021/acs.est.7b01210.

Huneeus, N. *et al.* (2011) 'Global dust model intercomparison in AeroCom phase I', *Atmospheric Chemistry and Physics*, 11(15), pp. 7781–7816. Available at: https://doi.org/10.5194/acp-11-7781-2011.

Hussein, T. *et al.* (2020) 'Particulate Matter Concentrations in a Middle Eastern City – An Insight to Sand and Dust Storm Episodes', *Aerosol and Air Quality Research*, 20(12), pp. 2780–2792. Available at: https://doi.org/10.4209/aaqr.2020.05.0195.

Inness, A. *et al.* (2019) 'The CAMS reanalysis of atmospheric composition', *Atmospheric Chemistry and Physics*, 19(6), pp. 3515–3556. Available at: https://doi.org/10.5194/acp-19-3515-2019.

Jaeglé, L. *et al.* (2011) 'Global distribution of sea salt aerosols: new constraints from in situ and remote sensing observations', *Atmospheric Chemistry and Physics*, 11(7), pp. 3137–3157. Available at: https://doi.org/10.5194/acp-11-3137-2011.

Jetter, J.J. *et al.* (2002) 'Characterization of emissions from burning incense', *The Science of the Total Environment*, 295(1–3), pp. 51–67. Available at: https://doi.org/10.1016/s0048-9697(02)00043-8.

Ji, X. *et al.* (2010) 'Characterization of particles emitted by incense burning in an experimental house', *Indoor Air*, 20(2), pp. 147–158. Available at: https://doi.org/10.1111/j.1600-0668.2009.00634.x.

Jickells, T. and Moore, C.M. (2015) 'The Importance of Atmospheric Deposition for Ocean Productivity', *Annual Review of Ecology, Evolution, and Systematics*, 46(1), pp. 481–501. Available at: https://doi.org/10.1146/annurev-ecolsys-112414-054118.

Jin, C. *et al.* (2022) 'Global validation and hybrid calibration of CAMS and MERRA-2 PM2.5 reanalysis products based on OpenAQ platform', *Atmospheric Environment*, 274, p. 118972. Available at: https://doi.org/10.1016/j.atmosenv.2022.118972.

Kar, J. *et al.* (2018) 'CALIPSO lidar calibration at 532 nm: Version 4 nighttime algorithm', *Atmospheric Measurement Techniques*, 11, pp. 1459–1479. Available at: https://doi.org/10.5194/amt-11-1459-2018.

Karagulian, F. *et al.* (2015) 'Contributions to cities' ambient particulate matter (PM): A systematic review of local source contributions at global level', *Atmospheric Environment*, 120, pp. 475–483. Available at: https://doi.org/10.1016/j.atmosenv.2015.08.087.

Kaskaoutis, D.G. *et al.* (2019) 'Analysis of intense dust storms over the eastern Mediterranean in March 2018: Impact on radiative forcing and Athens air quality', *Atmospheric Environment*, 209, pp. 23–39. Available at: https://doi.org/10.1016/j.atmosenv.2019.04.025.

Ke, G. *et al.* (2017) 'LightGBM: A Highly Efficient Gradient Boosting Decision Tree', in *Advances in Neural Information Processing Systems*. Curran Associates, Inc. Available at: https://proceedings.neurips.cc/paper/2017/hash/6449f44a102fde848669bdd9eb6b76fa-Abstract.html (Accessed: 9 December 2022).

Kelly, K.E. *et al.* (2017) 'Ambient and laboratory evaluation of a low-cost particulate matter sensor', *Environmental Pollution*, 221, pp. 491–500. Available at: https://doi.org/10.1016/j.envpol.2016.12.039.

Kidder, S.Q. (2015) 'SATELLITES AND SATELLITE REMOTE SENSING | Orbits', in G.R. North, J. Pyle, and F. Zhang (eds) *Encyclopedia of Atmospheric Sciences (Second Edition)*. Oxford: Academic Press, pp. 95–106. Available at: https://doi.org/10.1016/B978-0-12-382225-3.00362-5.

Kim, K.-H., Kabir, E. and Kabir, S. (2015) 'A review on the human health impact of airborne particulate matter', *Environment International*, 74, pp. 136–143. Available at: https://doi.org/10.1016/j.envint.2014.10.005.

Kim, M. *et al.* (2014) 'Improvement of aerosol optical depth retrieval over Hong Kong from a geostationary meteorological satellite using critical reflectance with background optical depth correction', *Remote Sensing of Environment*, 142, pp. 176–187. Available at: https://doi.org/10.1016/j.rse.2013.12.003.

Kim, M.-H. *et al.* (2018) 'The CALIPSO version 4 automated aerosol classification and lidar ratio selection algorithm', *Atmospheric Measurement Techniques*, 11(11), pp. 6107–6135. Available at: https://doi.org/10.5194/amt-11-6107-2018.

Kleinbaum, D.G. *et al.* (2013) *Applied Regression Analysis and Other Multivariable Methods*. Cengage Learning.

Kleist, D.T. *et al.* (2009) 'Introduction of the GSI into the NCEP Global Data Assimilation System', *Weather and Forecasting*, 24(6), pp. 1691–1705. Available at: https://doi.org/10.1175/2009WAF2222201.1.

Knippertz, P. and Todd, M.C. (2010) 'The central west Saharan dust hot spot and its relation to African easterly waves and extratropical disturbances', *Journal of Geophysical Research: Atmospheres*, 115(D12). Available at: https://doi.org/10.1029/2009JD012819.

Koster, R.D., Darmenov, A.S. and da Silva, A.M. (2015) *The Quick Fire Emissions Dataset (QFED): Documentation of Versions 2.1, 2.2 and 2.4*. NASA/TM-2015-104606 /Vol. 38. Available at: https://ntrs.nasa.gov/citations/20180005253 (Accessed: 17 January 2023).

Krewski, D. (2009) 'Evaluating the Effects of Ambient Air Pollution on Life Expectancy', *New England Journal of Medicine*, 360(4), pp. 413–415. Available at: https://doi.org/10.1056/NEJMe0809178.

Künzli, N. *et al.* (2005) 'Ambient Air Pollution and Atherosclerosis in Los Angeles', *Environmental Health Perspectives*, 113(2), pp. 201–206. Available at: https://doi.org/10.1289/ehp.7523.

Laffly, D. (2020) 'A Brief History of Remote Sensing', in *TORUS 2 – Toward an Open Resource Using Services*. John Wiley & Sons, Ltd, pp. 1–21. Available at: https://doi.org/10.1002/9781119720553.ch1.

Lee, K.H. *et al.* (2009) 'Atmospheric Aerosol Monitoring from Satellite Observations: A History of Three Decades', in Y.J. Kim et al. (eds) *Atmospheric and Biological Environmental Monitoring*.

Dordrecht: Springer Netherlands, pp. 13–38. Available at: https://doi.org/10.1007/978-1-4020-9674-7_2.

de Leeuw, G. *et al.* (2015) 'Evaluation of seven European aerosol optical depth retrieval algorithms for climate analysis', *Remote Sensing of Environment*, 162, pp. 295–315. Available at: https://doi.org/10.1016/j.rse.2013.04.023.

Lélé, M.I. and Lamb, P.J. (2010) 'Variability of the Intertropical Front (ITF) and Rainfall over the West African Sudan–Sahel Zone', *Journal of Climate*, 23(14), pp. 3984–4004. Available at: https://doi.org/10.1175/2010JCLI3277.1.

Levy, R.C., Remer, L.A. and Dubovik, O. (2007) 'Global aerosol optical properties and application to Moderate Resolution Imaging Spectroradiometer aerosol retrieval over land', *Journal of Geophysical Research: Atmospheres*, 112(D13). Available at: https://doi.org/10.1029/2006JD007815.

Li, L. *et al.* (2018) 'Retrieval of Daily PM2.5 Concentrations Using Nonlinear Methods: A Case Study of the Beijing–Tianjin–Hebei Region, China', *Remote Sensing*, 10(12), p. 2006. Available at: https://doi.org/10.3390/rs10122006.

Liang, S., Li, X. and Wang, J. (eds) (2012) 'Chapter 1 - A Systematic View of Remote Sensing', in *Advanced Remote Sensing*. Boston: Academic Press, pp. 1–31. Available at: https://doi.org/10.1016/B978-0-12-385954-9.00001-0.

Liaskoni, M. *et al.* (2023) 'Modelling the European wind-blown dust emissions and their impact on particulate matter (PM) concentrations', *Atmospheric Chemistry and Physics*, 23(6), pp. 3629–3654. Available at: https://doi.org/10.5194/acp-23-3629-2023.

LIFE IP HungAIRy (2019) *HungAIRy*. Available at: http://www.hungairy.hu/projekt.

Lin, L. *et al.* (2022) 'Global association between atmospheric particulate matter and obesity: A systematic review and meta-analysis', *Environmental Research*, 209, p. 112785. Available at: https://doi.org/10.1016/j.envres.2022.112785.

Liu, X. *et al.* (2020) 'Low-cost sensors as an alternative for long-term air quality monitoring', *Environmental Research*, 185, p. 109438. Available at: https://doi.org/10.1016/j.envres.2020.109438.

Liu, Y. *et al.* (2005) 'Estimating Ground-Level PM2.5 in the Eastern United States Using Satellite Remote Sensing', *Environmental Science & Technology*, 39(9), pp. 3269–3278. Available at: https://doi.org/10.1021/es049352m.

Liu, Y. *et al.* (2022) 'Study on Fugitive Dust Control Technologies of Agricultural Harvesting Machinery', *Agriculture*, 12(7), p. 1038. Available at: https://doi.org/10.3390/agriculture12071038.

Löffler, T., Burgdorf, J. and Klinkner, S. (2021) 'Orbit-raising strategies for cost-efficient access to lower Medium Earth Orbit without risking space debris', *Acta Astronautica*, 185, pp. 37–41. Available at: https://doi.org/10.1016/j.actaastro.2021.04.041.

Löndahl, J. *et al.* (2006) 'A set-up for field studies of respiratory tract deposition of fine and ultrafine particles in humans', *Journal of Aerosol Science*, 37(9), pp. 1152–1163. Available at: https://doi.org/10.1016/j.jaerosci.2005.11.004.

Lothon, M. *et al.* (2008) 'Observation of the Diurnal Cycle in the Low Troposphere of West Africa', *Monthly Weather Review*, 136(9), pp. 3477–3500. Available at: https://doi.org/10.1175/2008MWR2427.1.

Lücking, C., Colombo, C. and McInnes, C.R. (2012) 'A passive satellite deorbiting strategy for medium earth orbit using solar radiation pressure and the J2 effect', *Acta Astronautica*, 77, pp. 197–206. Available at: https://doi.org/10.1016/j.actaastro.2012.03.026.

Luo, H. and Han, Y. (2021) 'Impacts of the Saharan air layer on the physical properties of the Atlantic tropical cyclone cloud systems: 2003–2019', *Atmospheric Chemistry and Physics*, 21(19), pp. 15171–15184. Available at: https://doi.org/10.5194/acp-21-15171-2021.

Luo, N. *et al.* (2015) 'Improved aerosol retrieval algorithm using Landsat images and its application for PM10 monitoring over urban areas', *Atmospheric Research*, 153, pp. 264–275. Available at: https://doi.org/10.1016/j.atmosres.2014.08.012.

Lyamani, H., Olmo, F.J. and Alados-Arboledas, L. (2005) 'Saharan dust outbreak over southeastern Spain as detected by sun photometer', *Atmospheric Environment*, 39(38), pp. 7276–7284. Available at: https://doi.org/10.1016/j.atmosenv.2005.09.011.

Ma, J., Xu, J. and Qu, Y. (2020) 'Evaluation on the surface PM2.5 concentration over China mainland from NASA's MERRA-2', *Atmospheric Environment*, 237, p. 117666. Available at: https://doi.org/10.1016/j.atmosenv.2020.117666.

Mahowald, N. *et al.* (2014) 'The size distribution of desert dust aerosols and its impact on the Earth system', *Aeolian Research*, 15, pp. 53–71. Available at: https://doi.org/10.1016/j.aeolia.2013.09.002.

Maleki, H. *et al.* (2022) 'The impact of meteorological parameters on PM10 and visibility during the Middle Eastern dust storms', *Journal of Environmental Health Science and Engineering*, 20(1), pp. 495–507. Available at: https://doi.org/10.1007/s40201-022-00795-1.

Malings, C. *et al.* (2020) 'Fine particle mass monitoring with low-cost sensors: Corrections and long-term performance evaluation', *Aerosol Science and Technology*, 54(2), pp. 160–174. Available at: https://doi.org/10.1080/02786826.2019.1623863.

Malm, W.C. *et al.* (1994) 'Spatial and seasonal trends in particle concentration and optical extinction in the United States', *Journal of Geophysical Research: Atmospheres*, 99(D1), pp. 1347–1370. Available at: https://doi.org/10.1029/93JD02916.

Masiol, M. *et al.* (2014) 'Thirteen years of air pollution hourly monitoring in a large city: Potential sources, trends, cycles and effects of car-free days', *Science of The Total Environment*, 494–495, pp. 84–96. Available at: https://doi.org/10.1016/j.scitotenv.2014.06.122.

Mehmood, K. *et al.* (2022) 'Predicting the quality of air with machine learning approaches: Current research priorities and future perspectives', *Journal of Cleaner Production*, 379, p. 134656. Available at: https://doi.org/10.1016/j.jclepro.2022.134656.

Mei, D. *et al.* (2018) 'The influence of wind speed on airflow and fine particle transport within different building layouts of an industrial city', *Journal of the Air & Waste Management Association (1995)*, 68(10), pp. 1038–1050. Available at: https://doi.org/10.1080/10962247.2018.1465487.

Meo, S.A. *et al.* (2022) 'Impact of sandstorm on environmental pollutants PM2.5, carbon monoxide, nitrogen dioxide, ozone, and SARS-CoV-2 morbidity and mortality in kuwait', *Journal of King Saud University - Science*, 34(5), p. 102109. Available at: https://doi.org/10.1016/j.jksus.2022.102109.

Michalos, A.C. (ed.) (2014) *Encyclopedia of Quality of Life and Well-Being Research*. Dordrecht: Springer Netherlands. Available at: https://doi.org/10.1007/978-94-007-0753-5.

Minguillón, M., Viana, M. and Querol, X. (2013) *Particulate matter: Environmental monitoring & mitigation*, p. 177. Available at: https://doi.org/10.4155/9781909453135.

Mishchenko, M.I. and Geogdzhayev, I.V. (2007) 'Satellite remote sensing reveals regional tropospheric aerosol trends', *Optics Express*, 15(12), pp. 7423–7438. Available at: https://doi.org/10.1364/OE.15.007423.

Montenbruck, O. and Gill, E. (2000) *Satellite Orbits*. Berlin, Heidelberg: Springer. Available at: https://doi.org/10.1007/978-3-642-58351-3.

Montgomery, D.C., Peck, E.A. and Vining, G.G. (2020) *Introduction to linear regression analysis*. Sixth edition. Hoboken, New Jersey: Wiley (Wiley series in probability and statistics).

Oluwadairo, T. *et al.* (2022) 'Effects of aerosol particle size on the measurement of airborne PM2.5 with a low-cost particulate matter sensor (LCPMS) in a laboratory chamber', *Environmental Monitoring and Assessment*, 194(2), p. 56. Available at: https://doi.org/10.1007/s10661-021-09715-6.

Parkin, D.W., Delany, A. C. and Delany, Audrey Claire (1967) 'A search for airborne cosmic dust on Barbados', *Geochimica et Cosmochimica Acta*, 31(8), pp. 1311–1320. Available at: https://doi.org/10.1016/S0016-7037(67)80017-6.

Perrone, M.G. *et al.* (2018) 'Sources and geographic origin of particulate matter in urban areas of the Danube macro-region: The cases of Zagreb (Croatia), Budapest (Hungary) and Sofia (Bulgaria)', *Science of The Total Environment*, 619–620, pp. 1515–1529. Available at: https://doi.org/10.1016/j.scitotenv.2017.11.092.

Peuch, V.-H. *et al.* (2018) 'The Use of Satellite Data in the Copernicus Atmosphere Monitoring Service (Cams)', *IGARSS 2018 - 2018 IEEE International Geoscience and Remote Sensing Symposium*, pp. 1594–1596. Available at: https://doi.org/10.1109/IGARSS.2018.8518698.

Pey, J. *et al.* (2013) 'African dust outbreaks over the Mediterranean Basin during 2001–2011: PM_{10} concentrations, phenomenology and trends, and its relation with synoptic and mesoscale meteorology', *Atmospheric Chemistry and Physics*, 13(3), pp. 1395–1410. Available at: https://doi.org/10.5194/acp-13-1395-2013.

Plocoste, T. *et al.* (2022) 'Quantifying Spatio-Temporal Dynamics of African Dust Detection Threshold for PM10 Concentrations in the Caribbean Area Using Multiscale Decomposition',

Frontiers in Environmental Science, 10. Available at: https://www.frontiersin.org/articles/10.3389/fenvs.2022.907440 (Accessed: 30 January 2023).

Prospero, J.M. (1968) 'atmospheric dust studies on Barbados', *Bulletin of the American Meteorological Society*, 49(6), pp. 645–652. Available at: https://doi.org/10.1175/1520-0477-49.6.645.

Prospero, J.M. *et al.* (2020) 'Characterizing and Quantifying African Dust Transport and Deposition to South America: Implications for the Phosphorus Budget in the Amazon Basin', *Global Biogeochemical Cycles*, 34(9), p. e2020GB006536. Available at: https://doi.org/10.1029/2020GB006536.

Prospero, J.M. *et al.* (2021) 'The Discovery of African Dust Transport to the Western Hemisphere and the Saharan Air Layer: A History', *Bulletin of the American Meteorological Society*, 102(6), pp. E1239–E1260. Available at: https://doi.org/10.1175/BAMS-D-19-0309.1.

Provençal, S. *et al.* (2017) 'Evaluation of PM surface concentrations simulated by Version 1 of NASA's MERRA Aerosol Reanalysis over Europe', *Atmospheric Pollution Research*, 8(2), pp. 374–382. Available at: https://doi.org/10.1016/j.apr.2016.10.009.

Querol, X. *et al.* (2009) 'African dust contributions to mean ambient PM10 mass-levels across the Mediterranean Basin', *Atmospheric Environment*, 43(28), pp. 4266–4277. Available at: https://doi.org/10.1016/j.atmosenv.2009.06.013.

Randles, C.A. *et al.* (2017) 'The MERRA-2 Aerosol Reanalysis, 1980 Onward. Part I: System Description and Data Assimilation Evaluation', *Journal of Climate*, 30(17), pp. 6823–6850. Available at: https://doi.org/10.1175/JCLI-D-16-0609.1.

Ravi, S. *et al.* (2011) 'Aeolian Processes and the Biosphere', *Reviews of Geophysics*, 49(3). Available at: https://doi.org/10.1029/2010RG000328.

Raysoni, A.U. *et al.* (2023) 'A Review of Literature on the Usage of Low-Cost Sensors to Measure Particulate Matter', *Earth*, 4(1), pp. 168–186. Available at: https://doi.org/10.3390/earth4010009.

Remoundaki, E. *et al.* (2011) 'PM10 composition during an intense Saharan dust transport event over Athens (Greece)', *Science of The Total Environment*, 409(20), pp. 4361–4372. Available at: https://doi.org/10.1016/j.scitotenv.2011.06.026.

Roberts, A. and Knippertz, P. (2012) 'Haboobs: convectively generated dust storms in West Africa', *Weather*, 67(12), pp. 311–316. Available at: https://doi.org/10.1002/wea.1968.

Roland, S. (2018) *UBC ATSC113 Flying - BL obstacle Turbulence*. Available at: https://www.eoas.ubc.ca/courses/atsc113/flying/met_concepts/03-met_concepts/03f-BL_obstacle_wake/index.html (Accessed: 1 November 2023).

Rossi, A., Petit, A. and McKnight, D. (2020) 'Short-term space safety analysis of LEO constellations and clusters', *Acta Astronautica*, 175, pp. 476–483. Available at: https://doi.org/10.1016/j.actaastro.2020.06.016.

Rostási, Á. *et al.* (2022) 'Saharan Dust Deposition in Central Europe in 2016—A Representative Year of the Increased North African Dust Removal Over the Last Decade', *Frontiers in Earth*

Science, 10. Available at: https://www.frontiersin.org/articles/10.3389/feart.2022.869902 (Accessed: 14 November 2023).

Salvador, P. *et al.* (2019) 'Synergistic effect of the occurrence of African dust outbreaks on atmospheric pollutant levels in the Madrid metropolitan area', *Atmospheric Research*, 226, pp. 208–218. Available at: https://doi.org/10.1016/j.atmosres.2019.04.025.

Salvador, P. *et al.* (2022) 'Increasing atmospheric dust transport towards the western Mediterranean over 1948–2020', *npj Climate and Atmospheric Science*, 5(1), pp. 1–10. Available at: https://doi.org/10.1038/s41612-022-00256-4.

Sayahi, T., Butterfield, A. and Kelly, K.E. (2019) 'Long-term field evaluation of the Plantower PMS low-cost particulate matter sensors', *Environmental Pollution*, 245, pp. 932–940. Available at: https://doi.org/10.1016/j.envpol.2018.11.065.

Sayeed, A. *et al.* (2022) 'Hourly and Daily PM2.5 Estimations Using MERRA-2: A Machine Learning Approach', *Earth and Space Science*, 9(11), p. e2022EA002375. Available at: https://doi.org/10.1029/2022EA002375.

Sayer, A.M. *et al.* (2014) 'MODIS Collection 6 aerosol products: Comparison between Aqua's e-Deep Blue, Dark Target, and "merged" data sets, and usage recommendations', *Journal of Geophysical Research: Atmospheres*, 119(24), p. 13,965-13,989. Available at: https://doi.org/10.1002/2014JD022453.

Seinfeld, J.H. and Pandis, S.N. (2016) *Atmospheric Chemistry and Physics: From Air Pollution to Climate Change*. John Wiley & Sons.

Shaw, A.A. and Gopalan, N.P. (2014) 'Finding frequent trajectories by clustering and sequential pattern mining', *Journal of Traffic and Transportation Engineering (English Edition)*, 1(6), pp. 393–403. Available at: https://doi.org/10.1016/S2095-7564(15)30289-0.

Shepherd, G. *et al.* (2016) *Global Assessment of Sand and Dust Storms*.

Sheridan, R.P., Liaw, A. and Tudor, M. (2021) 'Light Gradient Boosting Machine as a Regression Method for Quantitative Structure-Activity Relationships'. Available at: https://doi.org/10.48550/ARXIV.2105.08626.

Shi, H. *et al.* (2023) 'Dynamic association of ambient air pollution with incidence and mortality of pulmonary hypertension: A multistate trajectory analysis', *Ecotoxicology and Environmental Safety*, 262, p. 115126. Available at: https://doi.org/10.1016/j.ecoenv.2023.115126.

Shin, M. *et al.* (2020) 'Estimating ground-level particulate matter concentrations using satellite-based data: a review', *GIScience & Remote Sensing*, 57(2), pp. 174–189. Available at: https://doi.org/10.1080/15481603.2019.1703288.

Sofer, T. *et al.* (2013) 'Exposure to airborne particulate matter is associated with methylation pattern in the asthma pathway', *Epigenomics*, 5(2), pp. 147–154. Available at: https://doi.org/10.2217/epi.13.16.

Solomos, S. *et al.* (2018) 'From Tropospheric Folding to Khamsin and Foehn Winds: How Atmospheric Dynamics Advanced a Record-Breaking Dust Episode in Crete', *Atmosphere*, 9(7), p. 240. Available at: https://doi.org/10.3390/atmos9070240.

Song, Z. *et al.* (2018) 'Diurnal and seasonal variability of PM2.5 and AOD in North China plain: Comparison of MERRA-2 products and ground measurements', *Atmospheric Environment*, 191, pp. 70–78. Available at: https://doi.org/10.1016/j.atmosenv.2018.08.012.

Stein, A.F. *et al.* (2015) 'NOAA's HYSPLIT Atmospheric Transport and Dispersion Modeling System', *Bulletin of the American Meteorological Society*, 96(12), pp. 2059–2077. Available at: https://doi.org/10.1175/BAMS-D-14-00110.1.

Sundram, T.K.M. *et al.* (2022) 'Effects of Ambient Particulate Matter (PM2.5) Exposure on Calorie Intake and Appetite of Outdoor Workers', *Nutrients*, 14(22), p. 4858. Available at: https://doi.org/10.3390/nu14224858.

Tecer, L.H. *et al.* (2008) 'Particulate Matter (PM2.5, PM10-2.5, and PM10) and Children's Hospital Admissions for Asthma and Respiratory Diseases: A Bidirectional Case-Crossover Study', *Journal of Toxicology and Environmental Health, Part A*, 71(8), pp. 512–520. Available at: https://doi.org/10.1080/15287390801907459.

Thunis, P. *et al.* (2018) 'PM2.5 source allocation in European cities: A SHERPA modelling study', *Atmospheric Environment*, 187, pp. 93–106. Available at: https://doi.org/10.1016/j.atmosenv.2018.05.062.

Touré, N.E., Konaré, A. and Silué, S. (2012) 'Intercontinental Transport and Climatic Impact of Saharan and Sahelian Dust', *Advances in Meteorology*, 2012, p. e157020. Available at: https://doi.org/10.1155/2012/157020.

Tran, L.K. *et al.* (2021) 'The impact of incense burning on indoor PM2.5 concentrations in residential houses in Hanoi, Vietnam', *Building and Environment*, 205, p. 108228. Available at: https://doi.org/10.1016/j.buildenv.2021.108228.

'Types of Satellite Orbits - mjginfologs.com' (2020), 5 December. Available at: https://mjginfologs.com/types-of-satellite-orbits/ (Accessed: 20 February 2023).

Ustin, S.L. and Middleton, E.M. (2021) 'Current and near-term advances in Earth observation for ecological applications', *Ecological Processes*, 10(1), p. 1. Available at: https://doi.org/10.1186/s13717-020-00255-4.

Uzunpinar, E.S. *et al.* (2023) 'Modification of Saharan dust size distribution during its transport over the Anatolian Plateau', *Science of The Total Environment*, 892, p. 164646. Available at: https://doi.org/10.1016/j.scitotenv.2023.164646.

Valavanidis, A., Fiotakis, K. and Vlachogianni, T. (2008) 'Airborne Particulate Matter and Human Health: Toxicological Assessment and Importance of Size and Composition of Particles for Oxidative Damage and Carcinogenic Mechanisms', *Journal of Environmental Science and Health, Part C*, 26(4), pp. 339–362. Available at: https://doi.org/10.1080/10590500802494538.

Valle-Di´az, C.J. *et al.* (2016) 'Impact of Long-Range Transported African Dust on Cloud Water Chemistry at a Tropical Montane Cloud Forest in Northeastern Puerto Rico', *Aerosol and Air Quality Research*, 16(3), pp. 653–664. Available at: https://doi.org/10.4209/aaqr.2015.05.0320.

Van Donkelaar, A., Martin, R.V. and Park, R.J. (2006) 'Estimating ground-level PM2.5 using aerosol optical depth determined from satellite remote sensing', *Journal of Geophysical Research: Atmospheres*, 111(D21). Available at: https://doi.org/10.1029/2005JD006996.

Varga, G. (2020) ‘Changing nature of Saharan dust deposition in the Carpathian Basin (Central Europe): 40 years of identified North African dust events (1979–2018)’, *Environment International*, 139, p. 105712. Available at: https://doi.org/10.1016/j.envint.2020.105712.

Varga, G., Kovács, J. and Újvári, G. (2013) ‘Analysis of Saharan dust intrusions into the Carpathian Basin (Central Europe) over the period of 1979–2011’, *Global and Planetary Change*, 100, pp. 333–342. Available at: https://doi.org/10.1016/j.gloplacha.2012.11.007.

Vlachokostas, Ch. *et al.* (2011) ‘Multicriteria methodological approach to manage urban air pollution’, *Atmospheric Environment*, 45(25), pp. 4160–4169. Available at: https://doi.org/10.1016/j.atmosenv.2011.05.020.

Wagner, A. *et al.* (2021) ‘Comprehensive evaluation of the Copernicus Atmosphere Monitoring Service (CAMS) reanalysis against independent observations’, *Elementa: Science of the Anthropocene*, 9(1), p. 00171. Available at: https://doi.org/10.1525/elementa.2020.00171.

Wang, Q., Gu, J. and Wang, X. (2020) ‘The impact of Sahara dust on air quality and public health in European countries’, *Atmospheric Environment*, 241, p. 117771. Available at: https://doi.org/10.1016/j.atmosenv.2020.117771.

Wang, Y. *et al.* (2015) ‘Laboratory Evaluation and Calibration of Three Low-Cost Particle Sensors for Particulate Matter Measurement’, *Aerosol Science and Technology*, 49(11), pp. 1063–1077. Available at: https://doi.org/10.1080/02786826.2015.1100710.

Wei, X. *et al.* (2020) ‘Satellite remote sensing of aerosol optical depth: advances, challenges, and perspectives’, *Critical Reviews in Environmental Science and Technology*, 50(16), pp. 1640–1725. Available at: https://doi.org/10.1080/10643389.2019.1665944.

Weidinger, T. *et al.* (2010) ‘Urban Air Quality, Long Term Trends and Road Traffic Air Pollution Modeling of Szeged’, in. Available at: https://doi.org/10.5772/10180.

Wen, Y. *et al.* (2022) ‘Quantitatively Disentangling the Geographical Impacts of Topography on PM2.5 Pollution in China’, *Remote Sensing*, 14(24), p. 6309. Available at: https://doi.org/10.3390/rs14246309.

Williams, R. *et al.* (2015) *Evaluation of field-deployed low cost PM sensors*. Technical EPA/600/R-14/464. Research Triangle Park, NC: U.S. Enviromental Protection Agency, Office of Research and Development, National Exposure Research Laboratory, p. 76.

Winker, D.M. *et al.* (2010) ‘The CALIPSO Mission: A Global 3D View of Aerosols and Clouds’, *Bulletin of the American Meteorological Society*, 91(9), pp. 1211–1230. Available at: https://doi.org/10.1175/2010BAMS3009.1.

World Health Organization (2021) *WHO global air quality guidelines: particulate matter (PM2.5 and PM10), ozone, nitrogen dioxide, sulfur dioxide and carbon monoxide*. Geneva: World Health Organization. Available at: https://apps.who.int/iris/handle/10665/345329 (Accessed: 7 April 2022).

Xia, Z. *et al.* (2021) *A Novel Integrated Network with LightGBM for Click-Through Rate Prediction*. Available at: https://doi.org/10.21203/rs.3.rs-872310/v1.

Yeganeh, B. *et al.* (2017) ‘A satellite-based model for estimating PM2.5 concentration in a sparsely populated environment using soft computing techniques’, *Environmental Modelling & Software*, 88, pp. 84–92. Available at: https://doi.org/10.1016/j.envsoft.2016.11.017.

Zhang, S. *et al.* (2020) ‘Review of aerosol optical depth retrieval using visibility data’, *Earth-Science Reviews*, 200, p. 102986. Available at: https://doi.org/10.1016/j.earscirev.2019.102986.

Zhang, X. *et al.* (2022) ‘Dust-planetary boundary layer interactions amplified by entrainment and advections’, *Atmospheric Research*, 278, p. 106359. Available at: https://doi.org/10.1016/j.atmosres.2022.106359.

Zhou, L. *et al.* (2021) ‘Rising Planetary Boundary Layer Height over the Sahara Desert and Arabian Peninsula in a Warming Climate’, *Journal of Climate*, 34(10), pp. 4043–4068. Available at: https://doi.org/10.1175/JCLI-D-20-0645.1.

8.2 Supplementary figures

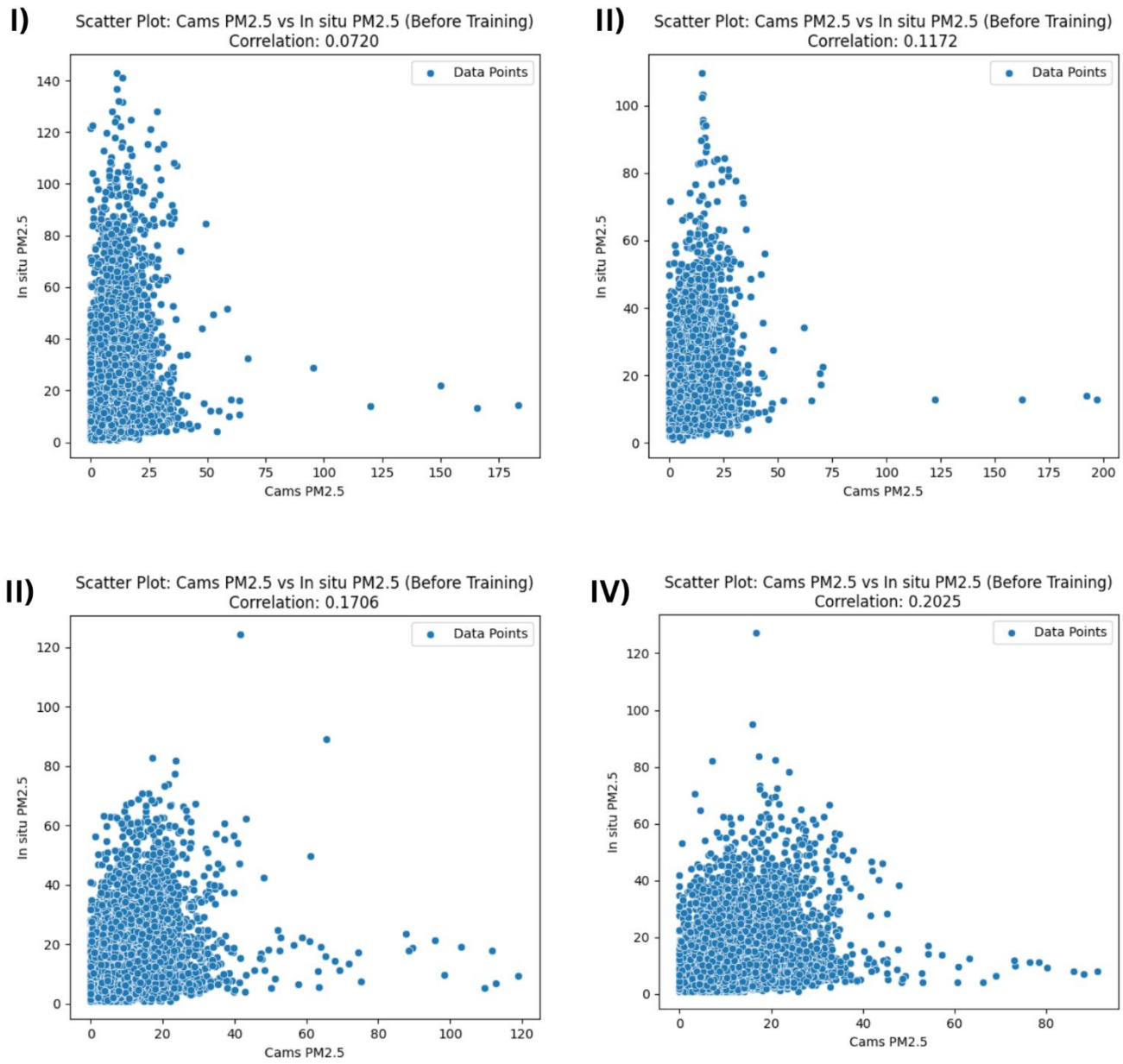

Figure 8.1: Scatter plots of raw CAMS PM2.5 data for I) Miskolc Alfoldi station, II) Nyiregyhaza, III) Pecs Nevelesi Kozpont station, IV) Szazhalombatta Buzavirag ter station

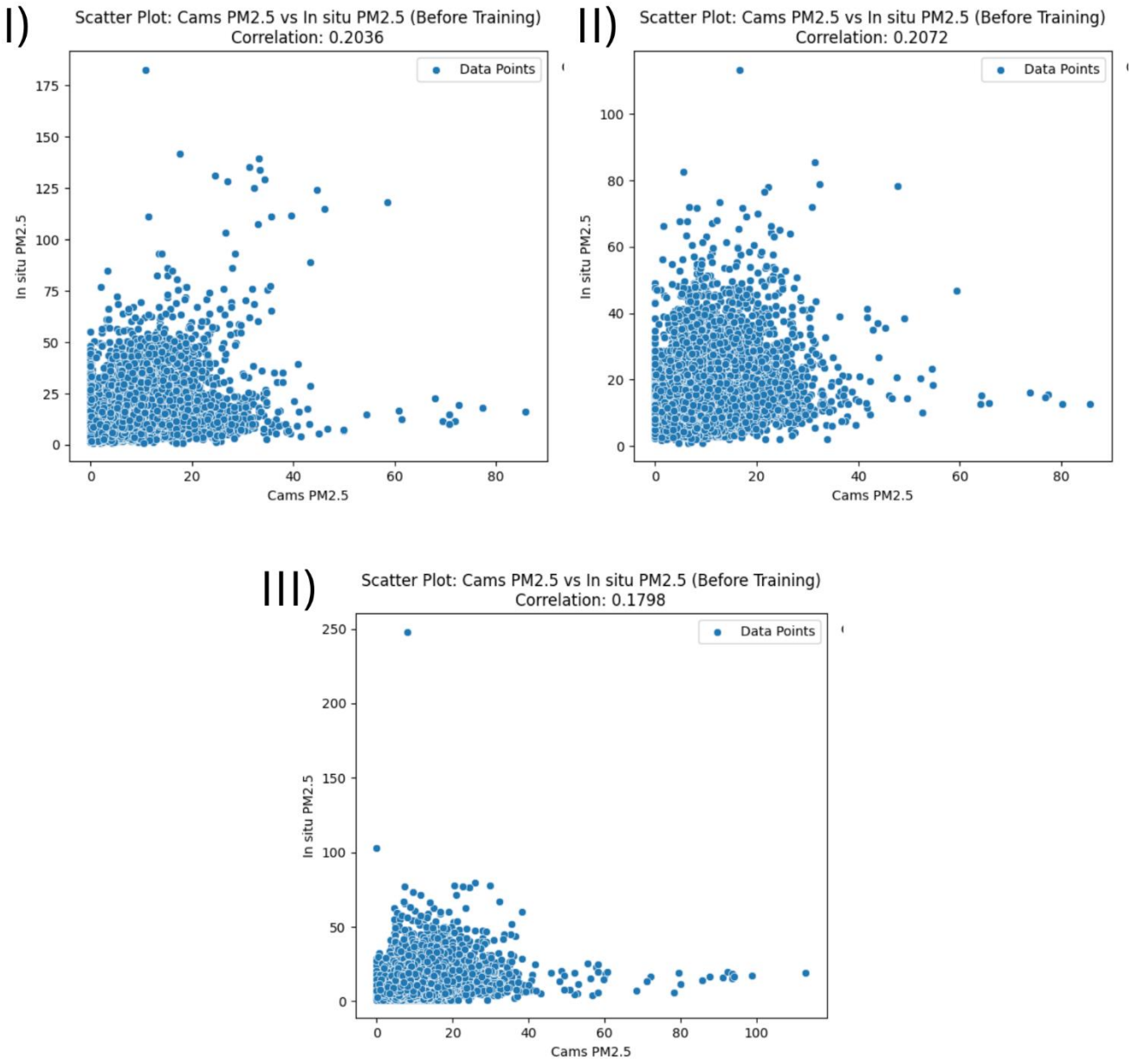

Figure 8.2: Scatter plots of raw CAMS PM2.5 data for I) Szeged 2 Rozsa station, II) Szolnok station, and III) Veszprem station

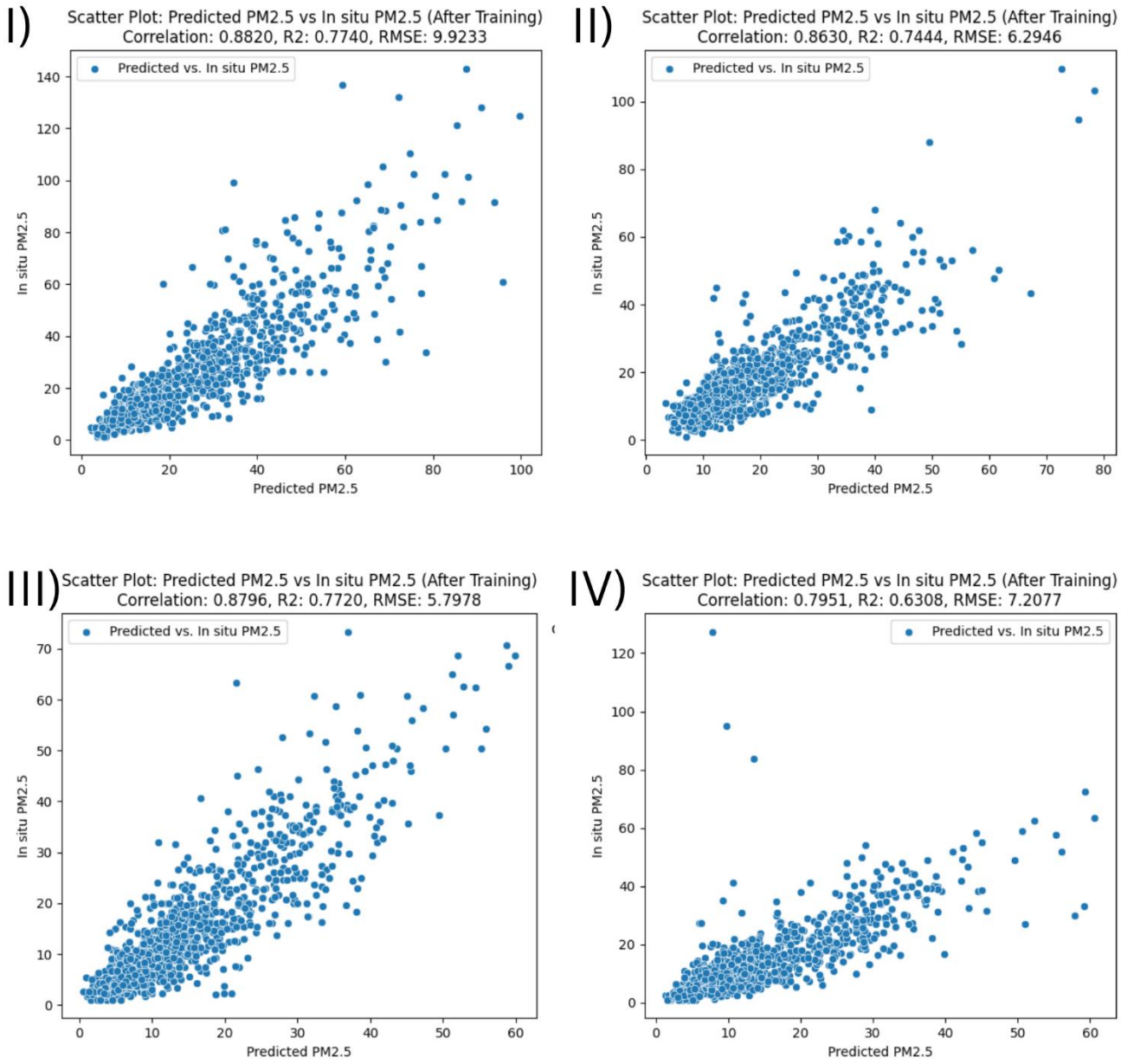

Figure 8.3: Scatter plots of predicted and In-situ PM2.5 for I) Miskolc Alfoldi station, II) Nyiregyhaza, III) Pecs Nevelesi Kozpont station, IV) Szazhalombatta Buzavirag ter station

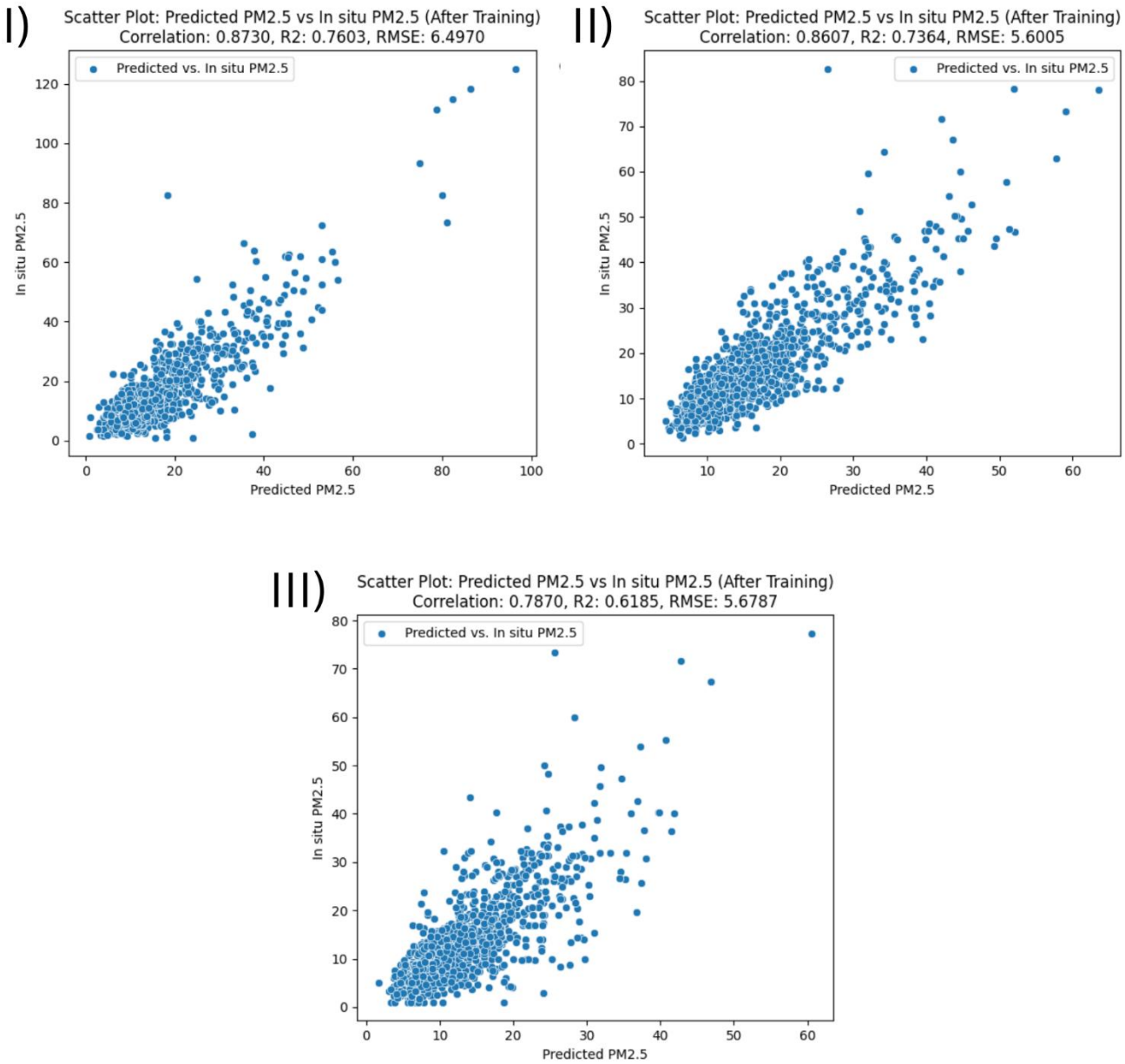

Figure 8.4: Scatter plots of predicted and In-situ PM2.5 I) Szeged 2 Rozsa station, II) Szolnok station, and III) Veszprem station