Air quality data analysis obtained by means of new generation sensor network

Nicoletta Lotrecchiano

UNIVERSITY OF SALERNO

DEPARTMENT OF INDUSTRIAL ENGINEERING

Ph.D. Course in Industrial Engineering
Curriculum in Chemical Engineering - XXXIV Cycle

AIR QUALITY DATA ANALYSIS OBTAINED BY MEANS OF NEW GENERATION SENSOR NETWORK

Supervisor
Prof. Massimo Poletto

Ph.D. student
Nicoletta Lotrecchiano

Scientific Referees
Prof. Diego Barletta
Ing. Daniele Sofia

Ph.D. Course Coordinator
Prof. Francesco Donsì

To Letizia and Lorenzo

List of published papers

Giuliano, A., Gioiella, F., Sofia, D., Lotrecchiano, N. (2018) A novel methodology and technology to promote the social acceptance of biomass power plants avoiding nimby syndrome. *Chem. Eng. Trans.*, **67**, 307-312, https://doi.org/10.3303/CET1867052.

Sofia, D., Lotrecchiano, N., Giuliano, A., Barletta, D., Poletto, M. (2019) Optimization of number and location of sampling points of an air quality monitoring network in an urban contest. *Chem. Eng. Trans*., **74**, 277–282, doi:10.3303/CET1974047.

Lotrecchiano, N., Gioiella, F., Giuliano, A., Sofia, D. (2019) Forecasting Model Validation of Particulate Air Pollution by Low Cost Sensors Data. *J. Model. Optim*., **11**, 63–68, doi:10.32732/jmo.2019.11.2.63.

Lotrecchiano, N., Sofia, D., Giuliano, A., Barletta, D., Poletto, M. (2019) Real-time on-road monitoring network of air quality. *Chem. Eng. Trans*., **74**, 241–246, doi:10.3303/CET1974041.

Lotrecchiano, N., Sofia, D., Giuliano, A., Barletta, D., Poletto, M. (2020) Pollution dispersion from a fire using a Gaussian plume model. *Int. J. Saf. Secur. Eng*., **10**, 431–439, doi:10.18280/ijsse.100401.

Sofia, D., Gioiella, F., Lotrecchiano, N., Giuliano, A. (2020) Mitigation strategies for reducing air pollution. *Environ. Sci. Pollut. Res*., **27,** 19226–19235, doi:10.1007/s11356-020-08647-x.

Sofia, D., Gioiella, F., Lotrecchiano, N., Giuliano, A. (2020) Cost-benefit analysis to support decarbonization scenario for 2030: A case study in Italy. *Energy Policy*, **137**, 111137, doi:10.1016/j.enpol.2019.111137.

Sofia, D., Lotrecchiano, N., Cirillo, D., La Villetta, M. (2020) NO2 Dispersion model of emissions of a 20 kwe biomass gasifier. *Chem. Eng. Trans*., **82**, 451–456, doi:10.3303/CET2082076.

Sofia, D., Lotrecchiano, N., Cirillo, D., La Villetta, M. (2020) Environmental Impact Analysis of Flue Gases Emissions for a 20 kWe Biomass Gasifier. *Chem. Eng. Trans*., **80**, 295-300, doi:10.3303/CET2080050

Sofia, D., Lotrecchiano, N., Trucillo, P., Giuliano, A., Terrone, L. (2020) Novel air pollution measurement system based on ethereum blockchain. *J. Sens. Actuator Netw*., **9**, 49, doi:10.3390/jsan9040049.

Lotrecchiano, N., Sofia, D., Giuliano, A., Barletta, D., Poletto, M. (2021) Spatial Interpolation Techniques For innovative Air Quality Monitoring Systems. *Chem. Eng. Trans*., **86**, 391–396, doi:10.3303/CET2186066.

Lotrecchiano, N., Capozzi, V., Sofia, D. (2021) An Innovative Approach to Determining the Contribution of Saharan Dust to Pollution. *Int. J. Environ. Res. Public Health*, **18**, 6100, doi:10.3390/ijerph18116100.

Lotrecchiano, N., Trucillo, P., Barletta, D., Poletto, M., Sofia, D. (2021) Air Pollution Analysis during the Lockdown on the City of Milan. *Processes*, **9**, 1692, https://doi.org/10.3390/pr9101692.

Lotrecchiano, N., Ciampa, R., Giuliano, A., Sofia, D. (2022). Study of the waste disposal effect on air quality: the Italian case study. *Int. Jour. of Env. and Sust. Dev*., **21**, 4, DOI:10.1504/IJESD.2021.10040655.

Lotrecchiano, N., Montano, L., Bonapace, I. M., Giancarlo, T., Trucillo, P., Sofia, D. (2022) Comparison Process of Blood Heavy Metals Absorption Linked to Measured Air Quality Data in Areas with High and Low Environmental Impact. *Processes*, **10**, 7, doi: 10.3390/pr10071409.

Outline

List of Figures

List of Tables

Table I.1. *Sensor characteristics* ,,,

Abstract

Air pollution measurement networks are meant to monitor the concentration of the primary pollutants. This type of infrastructures can be installed in urban areas in order to have an overall view of the environmental conditions. This thesis work aims to increase the quality of environmental data provided by air quality monitoring networks. In particular, the object of this thesis work is the treatment of the large number of data produced by the air quality monitoring networks implemented by Sense Square along the Italian territory both in urban and extra-urban context.

The first issue faced was the definition of the correct positions for the sampling points of the network that is relevant to the network design. A modelistic approach was applied to define the most interesting points of the city where to install the measuring stations. The Operational Street Pollution Model was used to simulate the PM10 concentrations in five chosen points. For each point it was possible to define the local conditions such as traffic intensity, street geometry, hourly number and type of vehicles passing through the streets and the meteorological parameters. The model adopted takes into account the pollutant dispersion using a Gaussian plume, while a box model is used to calculate the contribution from the recirculation vortex along the urban canyon. The state of art uses the dispersion models to define the pollutant behaviour dispersed in the atmosphere but doesn't compare the modelled results with the real measured ones. In this case, the comparison of the data obtained by the OSPM model and the ones measured after the installation of the sampling stations showed a good agreement, suggesting that the points chosen were the most interesting from the air pollution point of view.

A similar approach was used, to forecast the urban quality air using the concentration levels of particulate matter and meteorological conditions. The air quality data recorded by monitoring stations installed in Battipaglia, a southern city of Italy, during the whole month of August 2018 were considered. From 1st September to 10th November, a comparison was made between the predicted concentrations of PM2.5 and PM10 using the model with the measured concentrations by monitoring stations to fine-tune the

model. Finally, a correlation between air pollution and some meteorological factors (wind speed and direction, humidity) was investigated.

To integrate the fixed monitoring network to have a higher space resolution, the real-time on-road monitoring stations were implemented. The real-time on-road mobile monitoring network provides for the concentrations of pollutant in an innovative configuration. In fact, the data coming from vans equipped with the measuring devices moving along the streets of Milan makes it possible to aggregate the data or by districts or by cell. The district aggregation allows a macro-analysis of the data while the cell aggregation makes it possible a micro-analysis. For the micro-analysis, the whole investigated area was divided into square cells of 1km side to form a grid.

For each cell, the large amount of data available allows the analysis of the hourly, weekly and, monthly average. This new type of monitoring has to be compared with the traditional one represented by the Milan municipality air quality network. From the comparison between the two sets of data it was possible to notice the good agreement between them, suggesting that even if the measurement techniques are different, the results are reliable. Obviously, the availability of data from the dynamic network is strictly connected with the vans operating hours from 09:00 to 17:00. Moreover, the urban context is full of traffic restrictions like traffic limited zones, pedestrian pathways and parks that doesn't allows the vans passage. So, it is possible that some of the cells in the grid which the city is divided into have no data available. This missing values can be estimated by using a geospatial interpolation model. In particular, an Ordinary Kriging model was applied to one-day data in order to fill all the grid cells with estimated day averages of pollutant concentrations.

The methodology of Kriging is well known in literature but there are very few works that implemented it for large datasets as those treated in this case, due the calculative burden increasing more than linearly with the data set size. In future, the application of Kriging model will be improved by developing a suitable implementation also for large datasets. Firstly, a single pollutant will be considered and subsequently the model will be extended in order to include both the meteorological parameters and the traffic data that strongly influence the pollutant spatial dispersion.

Another possibility to estimate air quality in the missing cells, is to use measured data in the neighbouring cells with purposely designed algorithms.

In the present study, the results of the applications of few simple methods is assessed by comparing, in each cell where available data are present, the values predicted by the algorithm using the neighbouring data end the experimental value. Different criteria will be adopted and compared. In some of them the influence of the neighbouring cells on the prediction of the local concentration is based exclusively on the distance between the centre of the influencing cell and the centre of the tested cell, in others the direction and the speed of the wind are taken into consideration in the calculation

procedure. The approach proposed involves various innovative aspects, in fact this study use air quality data provided by an extremely innovative technology really and currently measured in an Italian city. Moreover, the model is applied to a very large area. Finally, the developed method for the spatial estimation of missing data was compared with other tecniques to evaluate the best solution.

During the first part of the work a deterministic approach was used to to investigate the behaviour of unexpected events like fires. This methodology was studied in deep to evaluate the dispersion of the pollutant generated after a fire occurred in a factory. The fire was described as an equivalent stack having the height of the observed cloud of smoke generated by the fire. The pollutant propagation was simulated with a Gaussian plume dispersion model. In particular, the unknown equivalent stack flow rate in the model was adjusted using the available data of PM2.5 on the ground at changing the air stability during night and day time and wind direction. The data available came from a fixed monitoring station located 4 km West-South-West from the fire. Provided the model flow rate best fitting the available data, the time and space distribution of PM2.5 concentrations at the ground level was estimated from the model.

From February 2020, the progressive adoption of measures to contain Coronavirus's contagion has resulted in Italy, especially in Lombardy, a sudden change in anthropic activities. From a scientific point of view, the new situation represents a unique laboratory for understanding and predicting the consequences of specific measures aimed at improving air quality. In this part of the work, the effect of the lockdown on air quality in the city of Milan (Italy) was analyzed. The PM10 values measured by the ARPA Lombardia air quality monitoring network indicate the seasonality of these pollutants, which typically record the highest values in the coldest months of the year. March 2020 data analysis shows an alternation of days with higher and lower particulate matter concentrations values. Some episodes highlighted the complexity of the phenomena related to the formation, transport, and accumulation of atmospheric particulates. Others highlighted the contribution of the second component and the meteorological situation most favorable to accumulation. The study showed that the trend of a general reduction of pollutant concentrations observed must be attributed to the decrease in emissions, in particular from the transport sector, from the variation of meteorological and environmental conditions.

Among the various phenomena that participate in air pollution, there is the influence of Saharan dust. These powders can participate in the increase of PM10 concentrations even though they are of a different nature compared to the particulate produced in urban areas. For this reason, a procedure has been implemented to take into account the contribution of these dusts to the extent of pollution.

Introduction

I.2.1 Air quality monitoring interest.

The environmental theme is one of the most discussed in the recent years. The attention of political groups, associations, and individual citizens is increasingly focused on the air quality. The efforts made to raise awareness of this issue are significant and such as to be able to put in place measures to improve air quality (Sofia et al., 2018a). Air pollution measurement networks are born to monitor the concentration of the primary pollutants. This type of infrastructures is installed in urban areas in order to have an overall view of the environmental conditions.

The effects of environmental pollution on human health are commonly known and people are concerned about the situation around them. In recent years, the so called NIMBY (Not In My Back-Yard) syndrome, referred to the public fear of the effects of possible pollution coming from waste treatment plants, has spread widely (Giuliano et al., 2018) and an accurate environmental policy can increase people awareness on the environment. In different industrial applications even if is not strictly required, is warmly recommended to monitor the particulate emissions in the working areas (Kahrizsangi et al., 2015; Sofia et al., 2013). Furthermore, health protection policies depend on data deriving from air quality monitoring surveys (Francis et al., 2008).

The major pollutants in urban and extra-urban environments are particulate matter (PM10 and PM2.5). Particulate Matter (PM) causes acute and chronic effects, particularly at the respiratory level since they can penetrate deep into the lungs. Primary PM sources are industrial production, transport and residential. Considerable relative humidity usually causes increases in PM concentrations due to the hygroscopic effect of aerosols, but not for PM10 in spring and summer, mainly due to the suppression of dust emissions under wet air conditions in spring and the impact of wet scavenging under high summer rainfall (Li et al., 2017).

Numerous epidemiological studies demonstrated the negative effects of exposure to high concentrations of particulate matter PM10 and PM2.5

(Brunekreef and Holgate, 2002), which cause an increase in the number of cardiovascular diseases (Dockery, 2001), pulmonary (Pope et al., 2002; deHartog et al., 2003), also in children (Buonanno et al., 2018), and a higher incidence in the number of deaths (Dockery et al., 1993; Burnett et al., 2000; Samet et al., 2000; Pope et al., 1995; Schwartz et al., 2002).

I.2.2 Monitoring campaigns.

Starting in the 2000s, monitoring campaigns were conducted in various cities around the world; one of the first was carried out in Izmir, Turkey (Elbir, 2003), with the aim of comparing the measured SO2 concentration values and those deriving from a forecasting model. This campaign was conducted with the help of four units which covered an area of 50x50 km.
Other cities such as Jamshedpur in India (Sivacoumar 2001), Tokyo (Tsujita et al., 2005), Pamplona (Parra et al., 2010), Helsinki and Athens (Vlachogianni et al., 2011), have been the subject of monitoring campaigns. using fixed stations for detecting the concentrations of the main pollutants such as PM2.5, PM10, CO, SO2, NO2, C6H6. Most of the environmental literature uses data coming from air quality monitoring networks to carry out firstly a Principal Component Analysis (PCA) to understand what are the parameters that mainly affects the phenomenon. Moreover, the data are used to study the seasonal variability of pollution.

The technology for the determination of the concentrations of pollutants dispersed in the air, after the modelling of the zone of installation, is turning towards the implementation of increasingly efficient and low-cost monitoring systems (Cariou et al., 2016) which can be easily installed in the city's hot spots. In fact, the low cost allows the installation of a higher number of monitoring stations, creating a high spatial resolution network, for a detailed knowledge of the local air conditions (Sofia et al., 2018b).

I.2.3 Mobile monitoring networks.

Some of the major manufacturers of monitoring networks have developed portable devices in recent years. Their small size allows their direct use by citizens, measuring all the polluting parameters supported by the device. However, none of these ready-to-use devices have efficient real-time detection systems or high performance (Thompson, 2016).
Recent studies have allowed the development of mobile monitoring systems in real time such as platforms active on vehicles. Monitoring systems directly on the road, such as those used for monitoring through the Netherlands (Weijers et al., 2004) and Zurich, Switzerland (Bukowiecki et al., 2002a, b, 2003), demonstrated how this measurement system can be a valid tool for assessing the spatial variability of the pollutant concentration. Moreover, it allows to investigate specific areas that it would not be possible

to achieve with fixed monitoring systems. By increasing the temporal resolution of these systems, it is also possible to improve the reliability of the measurements and the quantity of data collected. The devices used on the road as a mobile support of the monitoring station are usually automobiles (Devarakonda et al., 2013), specially equipped vehicles (Pirjola et al., 2004) or even alternative vehicles specifically designed for road monitoring, such as the bicycle Aeroflex (Elen et al., 2013). The advantage of mobile devices is their small size, which allows non-invasive installation even on small vehicles. A monitoring campaign could be initiated by installing mobile devices on the line buses or cars destined for car-sharing, so as to create a dense network of sensors that can well describe the situation in the city.

Real-time On-road Monitoring stations (ROMs) are a new kind of air quality monitoring units designed to be mounted on a carrying vehicle in motion within the urban area, in order to catch information on pollutant concentrations distributed in space and time. Using a limited number of devices mounted on vehicles moving along established tracks, it is possible to collect space and time distribution of pollutants otherwise unattainable. In this case, increasing the time resolution of the sensors allows to improve the reliability of measurements and the amount of data collected.

In recent years, many approaches are developed to predict air quality based on existing historical air quality and meteorological data. A model is a simplified representation of the reality and it gives an approximate description of the modeled phenomenon. One of the main purposes of modeling is the phenomenon explanation, sometimes it can be used to describe the mechanism behind the reality we are investigating. The last work on mobile sensing is by Mihăiţră et al. (2019) with the purpose of comparing the new technology represented by the mobile monitoring stations with the fixed sensing devices used by citizens to measure pollution. The mobile sensing device is represented by passive tubes for monitoring NO_2 concentration placed in strategic locations highly affected to traffic. The nitrogen dioxide was also monitored by citizens using smart and mobile pollution units carried at breathing level. The two pollution measuring methods were compared to reveal the difference between them. In the end the work applied a machine learning modelling by using decision trees and neural networks on the mobile-generated data and show that humidity and noise are the most important factors influencing the prediction of nitrogen dioxide concentrations of mobile stations.

The paper by Shiva Negendra SM et al. (2019) deals with the development of a smart personal monitoring system for real time air quality sensing to measure the individual exposure to air pollution. The low-cost and light-weight sensors (SPAMS) located on a bus (travelling from November 2015 to January 2016) were used to measure CO, NO_2, O_3, PM, temperature and humidity. The parameters were measured by walking on both footpaths and travelling in the bus during various times of a day (morning, afternoon and

evening) and different days in a week at selected locations in Chennai city, India. Lim et al. (2019) developed LUR models for street-level fine particulate matter (PM2.5) concentration levels in Seoul, South Korea. 169 h of data were collected from an approximately three weeks long campaign across five routes by ten volunteers sharing seven AirBeams, a low-cost smartphone-based particle counter, while geospatial data were extracted from OpenStreetMap, an open-source and crowd-generated geographical dataset. The five routes, four of which were based near or around government-run regulatory monitors, were designed to span various neighborhoods and to obtain spatial coverage of a wide range of types of geographical variables, such as major roads and highways, green spaces, and both low and high density residential areas.

The work carried out by Wen et al. (2019) reports an absolute principal component score (APCS) analysis of on-road mobile measurements to on-highway measurements of nitrogen oxides, carbon monoxide, carbon dioxide, black carbon, and particle number obtained from a mobile platform deployed over a 5-day sampling period in Chengdu, China. Data were collected for heavy-duty diesel truck (HDDT) plumes the general on-road environment. Pereira et al. (2018) proposed a student project which turned smart-phones into dynamic sensor nodes which send data to a centralized platform. The pollutant measured were NO_2, CO and, O_3 for the proposed solution that has an out-of-the-box central platform available in the cloud which gathers available IoT sensor data (ThingSpeak) and an online custom developed central platform. The study of Minet et al. (2015) highlights the challenges of short-term mobile sampling campaigns conducted in 2015 in Montreal. The study used NO_2 and O_3 Aeroqual Series 500 portable sensors, two types of GPS Garmin Edge 800 as well as MapMyRide a smartphone application. NO_2 and O_3 concentrations were recorded by cyclists and pedestrians. Hourly temperature, relative humidity and wind speed from the Montreal Pierre Elliot Trudeau International Airport weather station were recorded and synchronized with the other measurements. The work of Hasenfratz et al. (2015) analyze one of the largest spatially resolved ultrafine particles (UPF) data set publicly available which contains over 50 million air measurements system. The mobile measurement system consists of 10 sensor nodes installed on top of public transport vehicles for 2 years starting from April 2012, which are equipped with a semiconductor based O3 sensor, electrochemical-based CO and NO_2 sensors, and a compact device to measure UPF concentrations. Additionally, the nodes monitor radio-frequency electromagnetic fields, temperature and humidity. Van den Bossche et al. (2015) studied the impact of the temporal variability of a mobile measurement device and proposed a methodology to map urban air quality using mobile monitoring. A large set of black carbon measurements was collected in Antwerp, Belgium, using a bicycle equipped with a portable BC monitor (micro-aethalometer). The campaign consisted of 256 and 96

runs along two fixed routes (2 and 5 km long). Elen et al. (2012) presented the Aeroflex, a bicycle for mobile air quality monitoring. The Aeroflex was equipped with compact air quality measurement devices to monitor ultrafine particle number counts, particulate mass and black carbon concentrations at a high resolution (up to 1 second). Each measurement is automatically linked to its geographical location and time of acquisition using GPS and Internet time. Furthermore, the Aeroflex was equipped with automated data transmission, data pre-processing and data visualization. Over the past years, the Aeroflex was been successfully used for high resolution air quality mapping, exposure assessment and hot spot identification. The vehicular-based mobile approach for measuring fine –grained air quality in real time proposed by Devarakonda et al. (2013) was developed using two cost effective data farming models for the public transportation and a personal sensing device. Authors used a Mobile Sensing Box which can be mounted on vehicles and contains a micro-controller, dust and carbon monoxide sensors, GPS and a cellular modem. The personal sensing devices were represented by a mobile air quality sensor and a smartphone to act as an interface with the central repository hosted on a cloud server. They selected OXA and CLIMA modules to measure carbon monoxide, humidity, temperature, ambient light and barometric pressure. The work of Hachè (2011) introduces a novel wearable mobility monitoring system (WMMS) for an objective ubiquitous measurement of mobility. The WMMS prototype was created using a smartphone-based approach that allowed for an all-in-one WMMS. The wearable system is freely worn on a person's belt, such as a normal phone. The WMMS was designed to monitor a user's mobility state and to take a photograph when a change of state was detected. Al-Ali et al. (2010) proposed an online GPRS-Sensors Array for air pollution monitoring which consists of a Mobile Data-Acquisition Unit Mobile-DAQ) and a fixed Internet-Enabled Pollution Monitoring Server. The DAQ unit integrates a single chip microcontroller, air pollution sensors array, a GPRS-Modem, and a GPS-Module. The mobile measuring device was mounted on a university bus that was driven around the campus of the American University of Sharjah (AUS) to collect CO, NO_2, and SO_2 data for 12h every day. Pirjola et al. (2004) designed and built a mobile laboratory, "Sniffer", in Helsinki Polytechnic to measure traffic pollutants with high temporal and spatial resolution under real conditions. The van equipment's provides gas phase measurement of CO and NO_2, number size distribution measurements of fine and ultrafine particles with an electrical low pressure impactor, an ultrafine condensation particle counter and scanning mobility particle sizer. Also, meteorological and geographical parameters are recorded. The paper introduces the construction and technical details of the van, and presents data from the measurements performed during an LIPIKA campaign on the highway in Helsinki.

I.2.4 Models types.

The Joint Research Center (JRC) guidelines for the air quality monitoring, express the need to combine the normal techniques for detecting and measuring air quality with both meteorological models and models for calculating the pollutants dispersion (European Parliament, 2008). The impossibility, more often than not, of adequate space-time coverage by the monitoring stations, make the environmental parameters measurements alone insufficient to verify compliance with the limits imposed on existing sources or air quality values. It therefore becomes necessary to adopt additional tools capable of filling existing gaps and integrating the systems already operating in the area; such tools are represented by mathematical models. The high detail with which the models consider the spatial-temporal variations of emissions and meteorological conditions and the ability to assess the impact of emission sources and distinguish their contributions, make them useful in identifying effective environmental strategies.

The EPA - the American Agency for the Protection of the Environment - aware of the advantages that models can offer, identifies the best models on the market, tests their validity and effectiveness, updating and developing them continuously. AERMOD, CALPUFF, CALINE, ISC3, are some of the most used models that the American agency provides (EPA, 1995-2000).
The way of describing reality differentiates the models from each other and it is of utmost importance to understand which models are most suitable for simulating a given situation, what their reliability is and whether there is the possibility of implementation and improvement. A model is a tool capable of representing reality, used to predict and describe the evolution of a real phenomenon. The main requirements of a model are based on the fact that it must be able to predict the phenomenon trend taking into account any perturbations that induced it, include all a priori knowledge and be consistent with the theory that governs its construction. We speak of a mathematical model whenever any mathematical object is used to formalize the relationships existing between the variables that describe a real phenomenon. Mathematical models are divided into two classes: deterministic models and stochastic models.

In the field of atmospheric pollution, deterministic models are based on physical cause-effect relationships, while stochastic models are based on previous measurements at established points and only for those points it is possible to predict the future value of the concentration of pollutants. Stochastic models are used in cases where measurements are available and, therefore, generally in urban or industrial contexts, where real-time control of attention and alarm levels is essential (Artzrouni, 2006). Deterministic models are the most used; in them, the input variables assume fixed values, and the results obtained do not take into account any uncertainty factors (unlike the stochastic models). The deterministic models aim to

quantitatively reconstruct phenomena that determine the space-time evolution of the concentration of pollutants in the air. According to the different way of observing and describing the properties of a fluid, two classes of deterministic models can be distinguished: Eulerian models and Lagrangian models.

Eulerian models refer to a fixed coordinate system and are based on the integration of the differential diffusion Equation, obtained from the mass balance applied to an infinitesimal volume of air under certain assumptions. In essence, the Eulerian approach provides a fluid-dynamic description of a continuous medium and presupposes a totally deterministic view of the phenomenon. Depending on how the differential Equation is solved, analytical models (puff and Gaussian), box models and grid models will be obtained (Moussiopoulos et al., 1996).

The Gaussian analytical models are able to describe the downwind concentration on the ground due to a continuous point source, assuming a Gaussian law for the dispersion. They are easy to use models, because they adopt a series of restrictive simplifications: stationarity and homogeneity of weather conditions, horizontal wind speed not zero in the direction of the wind and on average zero on the plane orthogonal to it, flat terrain, absence of chemical transformations. The adoption of a plume-type dispersion model represents the emission of a stationary and continuous source.

The puff models represent an extension of the Gaussian models and allow to reconstruct the concentration values in non-homogeneous and non-stationary conditions. The value of the concentration in a point is obtained by adding the contributions of the concentrations of the various puffs within the domain, remembering that the dispersion for each puff always follows a Gaussian law. The adoption of a puff-type dispersion model represents the emission of a turbulent and discontinuous source. The box models divide the survey domain into one or more cells in which the pollutants are considered perfectly mixed. This allows you to rewrite the mass balance Equation in an easily solvable way and, known some parameters such as background concentration, wind speed, deposition speed, height of the mixing layer, allows you to easily calculate the concentration (Mareddy, 2018).

Grid models such as box models divide the computation domain into three-dimensional cells within which the solution of the atmospheric diffusion Equation is obtained through finite difference techniques, returning the concentration value for each point of the grid. As the number of nodes increases, the computing complexity increases; generally, the horizontal dimension of the cells is a few kilometers, while the vertical one is a function of the atmospheric layers that must be studied and can vary from a few meters to hundreds of kilometers.

Lagrangian models refer to a mobile coordinate system that follows the movements of the air masses whose behavior is to be reproduced. In the Lagrangian description, the pollutant is not considered a continuous fluid

introduced into another continuous fluid (air), but a set of particles (with characteristics similar to the concept of a particle given in Meteorology) independent of each other and identifiable individually, introduced into a fluid whose motion is known (mean motion and turbulent fluctuations). The description of the motion of each particle is therefore closely linked to the general motion of the fluid, it can no longer be considered deterministic, but stochastic due to the turbulence of the medium. Among them we can distinguish trajectory and particle models. In trajectory models, the evolution of a moving column of air is simulated under the action of the average component of the wind speed (by hypothesis horizontal and uniform with the altitude). In particle models the emission of pollutants is simulated with the generation of a certain number of particles emitted at each new time step: the concentration field at each step is reconstructed as a function directly proportional to the number of particles passing through a certain volume of space. Deterministic models describe the phenomena behavior of a real system, using algorithms capable of schematizing the system in question (Moussiopoulos et al., 1996).

General Equation of transport and diffusion

The Equation obtained by means of the mass balance, on an infinitesimal volume (*dxdydz*), during the time interval dt in its complete form is (Socolofsky, 2002):

$$\frac{\partial C}{\partial t} = -\nabla(vC) + \nabla(K * \nabla C) + \nabla(D * \nabla C) - R + S \qquad \text{(I.1)}$$

where:

- C is the dispersed pollutant concentration, [g/m^3];
- V is the vector of the wind intensity referred to the the reference heigth, [m/s];
- D are the molecular diffusivity coefficients, [m^2/s];
- K is the turbulence diffusivity tensor, [m^2/s];
- R is the removal factor, [g/(m^3·s)];
- S is the source [g/(m^3·s)].

The turbulent diffusivity tensor K is assumed diagonal (Kii = 0, Kij = 0) The removal factor R takes into account both the chemical removal component in the atmosphere and the removal component due to the deposition phenomenon. In practice, simplifying hypotheses are made in order to integrate eq. (I.1) and obtain solutions to the problem:

- Negligible molecular diffusion compared to turbulence.
- Negligible vertical component of wind speed ($\bar{w} = 0$). This hypothesis is not admissible in the case of sea breezes or in the presence of complex morphologies.

- Horizontal turbulent diffusivity coefficients Kxx and Kyy independent of x and y, while Kzz depends on the vertical coordinate z. In some applications Kxx and Kyy are neglected.
- Removal term linked to the chemistry of the atmosphere, negligible if the pollutant is inert or not very reactive.

According to the assumptions made so far, eq. (I.1) can be simplified as follows:

$$\frac{\partial C}{\partial t} + \bar{u}\frac{\partial C}{\partial x} + \bar{v}\frac{\partial C}{\partial y} = K_{xx}\frac{\partial^2 C}{\partial x^2} + K_{yy}\frac{\partial^2 C}{\partial y^2} + \left(\frac{\partial}{\partial z}\left(K_{zz}\frac{\partial^2 C}{\partial z}\right)\right) + s(x,y,z,t) \quad \text{(I.2)}$$

Eq. (I.2) contains a term that describes transport and a term that expresses diffusion. The dispersion of pollutants into the atmosphere, except for external contributions, is therefore mainly due to two processes: transport by the wind field and turbulent diffusion depending on the type of atmosphere. The analytical models solve eq. (I.2) analytically while the grid models do it through numerical calculation techniques.

The deposition

The term deposition indicates all those processes by which pollutants are transferred from the atmosphere to the earth's surface, depending on the type of receiving soil, the weather conditions, the pollutant considered and its concentration value in the atmosphere. The deposition phenomena are divided into dry and wet deposition; the latter being associated with precipitation events.
Dry deposition is a continuous phenomenon described by the defined deposition rate Like (Elimelech at al., 1998,):

$$v_d = \frac{F}{c(z)} \quad \text{(I.3)}$$

where:

v_d = deposition intensity, [m/s]
F = pollutant removed flux, [g/(m^2/s)]
c(z) = pollutant concentration near the ground, [g/m^3]

Dry deposition occurs in two phases, the first represents the pollutants transfer to a surface and the second is the pollution retention on that surface. The transfer is conditioned by aerodynamic effects responsible for transport near the surface by turbulent motions and depend on wind speed, atmospheric instability, the characteristics of the pollutant and the type of soil. The transfer is also conditioned by phoretic effects such as thermophoresis for which the particle moves, in the presence of a thermal

gradient, towards the lower temperature zone, diffusion-phoresis for which a particle is influenced by the concentration gradient and electrophoresis in which the electric fields of the medium or particle condition its motion. The dry deposition is also influenced by the effects regulated by the theory of the quasi laminar layer so that in the vicinity of the surface, turbulent motions no longer govern the course of pollutants. For gases we speak of molecular motions linked to viscous forces while with particles of size <0.1μm we speak of Brownian motions, and of inertia forces for larger diameters. The retention of pollutants by a surface basically depends on gases, chemical-biological properties of the surface itself; this absorbs, dissolves or involves pollutants in chemical reactions as appropriate. Particulate matter, on the other hand, can undergo resuspension and redeposition phenomena, this depends on both the type of soil, the type of pollutant and the wind speed. Associated with the deposition speed, we find the resistance to deposition expressed as:

$$r_d = \frac{1}{v_d} \tag{I.4}$$

where:

r_d = deposition resistance, [s/m].

The deposition resistance can be expressed as the sum of different resistances participating in the phenomenon. When considering turbulent transport resistance, resistance to transport through the quasi-laminar layer, resistance to surface retention, the total resistance can be calculated as the sum of the total resistance to transport through the atmospheric medium and the total resistance to transport through the atmospheric medium (eq. I.5).

$$r_d = r_a + r_b + r_s = r_{air} + r_s \tag{I.5}$$

Wet deposition includes all processes by which air pollutants are transported to the ground, through various forms of precipitation (rain, snow, fog). Two main processes can be distinguished: the rainout, in which the pollutant inside the clouds becomes the condensation core of the water drops and the washout, in which the pollutant below the clouds, hitting the water drops. water is dragged to the ground; in both cases, the water-pollutant interaction characterizes the process. In the washout process, the size of the water droplets, the PH of the rains and the type of pollutant species are some of the aspects to be evaluated. If we consider the case (aerosol or highly soluble gas) for which the leaching of pollutants by impact is an irreversible phenomenon, the law that describes the decrease in pollutants over time is of an exponential type:

$$Q(t + dt) = Q(t)e^{-I_w dt} \tag{I.6}$$

where:

Q(t) is the instantaneous pollutant at time t

l_w is the washout coefficient.

The washout coefficient is function of the rain intensity P(t), [mm/h] and its value is:

$$I_w = I_{w0} P(t)^a \quad \text{(I.7)}$$

l_{wo} is the constant that depends from the pollutant type while the exponential part a, assumes generally the value in the. range (0,75÷1).

Downwash phenomena

The presence of buildings near of a source can affect the dispersion of the emitted pollutant mass. The wake of buildings within a wind field generates an increase in mechanical turbulence and therefore an increase in total turbulence (Fig. I.1). In Figure I.1 it is possible to observe an area, called cavity, in the leeward part of the building, where it is assumed that the flows, due to the high turbulence, are completely mixed. The algorithms are not able to predict the concentration of pollutant in the cavity, so the calculation is performed at such distances as not to be affected by the influence of this (Canepa, 2004).

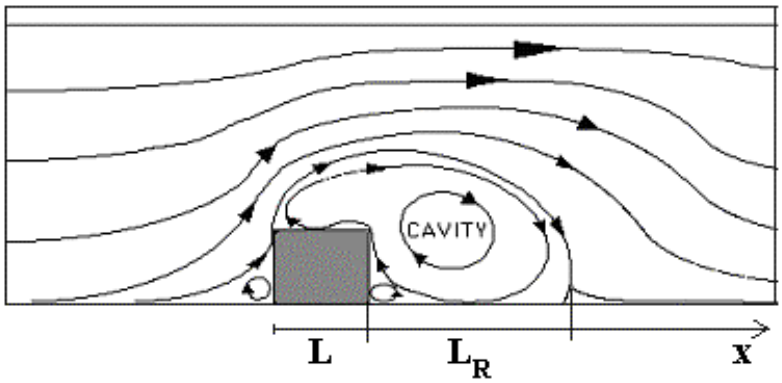


Figure I.1 *Trend of flow lines influenced by the presence of a building.*

The algorithms that describe downwash phenomena are activated only if the source is in a well-defined area near the building, both upwind and downwind of it. In addition, the guidelines provided by the EPA (1985): Guidline for Determination of Good Engineering Practice Stack Height (GEP) are used to determine whether the presence of buildings actually affects the plume. In these conditions, if the chimney is upwind or inside the "cavity" area, the flow lines are strongly distorted, with consequent rising of the plume and capturing part of it inside the "cavity", where it is created a

strong recirculation that homogenizes the concentrations on the vertical. Three types of concentration can be identified:

- CN: "near wake" concentration due to dilution caused by turbulence within the "cavity"; it is constant on the vertical and a function of the portion of the plume captured by the “cavity”.
- CF: concentration due to the part of the plume captured, considered as a source of volume.
- CP: concentration due to the part of the plume not captured, calculated as in the absence of downwash.

The calculation of the ground concentration, downwind of the building, is then carried out as a function of the distance from the building.
The most used deterministic dispersion models to define the pollutant dispersion are the Gaussian Plume and Puff dispersion that will be deeply described in the following chapters and applied to a case study.

Lagrangian models

As previously mentioned, Lagrangian models consider the fluid as the movement of a large number of particles in which the mass of each particle is conserved over time if no chemical or decay reactions take place. So the field of average concentration changes only for the redistribution of particles in space. The Lagrangian view simply follows the evolution over time of every single particle introduced into the fluid in turbulent motion. If the x axis is directed along the direction towards which the wind blows, the y axis in the transverse direction and z in the vertical direction, at an instant t0 a particle is emitted from a source located at the point $P_0 = P (x_0, y_0, z_0; t_0)$ and moves following the air movements. The emitted particle will move in an apparently disordered way, like air, and its motion can only be described in probabilistic terms. The Lagrangian theory therefore abandons a deterministic description of the phenomenon of transport and dispersion of a species in the air in favor of a purely statistics. A particle that at t_0 is in position $X_0 = (x_0, y_0, z_0)$, in the following instants t can move in the air with a disordered and random movement, reaching a new position X (x, y, z). Considering many different particles, all emitted at the point $X_0 = (x_0, y_0, z_0)$ at time t_0, at time t, due to the intrinsic randomness of their motion, deriving from the turbulence of the system, they may be in different points of the space, i.e. each particle will have a different trajectory. If we consider a point P (x, y, z) at a given instant t, the Lagrangian approach constructs a stochastic model of the distribution, of the particles emitted at time t, of the particles emitted at all previous instants, determining their influence in the desired point of space-time. The formal complexity of the general formulation of the Lagrangian model is remarkable. The basic stochastic entity for the Lagrangian model description is the Probability Density

Function (PDF) of a particle ψ (x, y, z; t) which represents the probability that a particle is at time t in a small volume of air centered on the point (x, y, z) of dimensions dx × dy × dz. Totally general considerations lead to a probabilistic definition of the pollutant concentration (and therefore of particles) in space (Moussiopoulos et al., 1996). If we assume that particles, each with mass qk, are present at time t in space M, in probabilistic terms the concentration can be expressed as:

$$< c(x, y, z; t) \geq \sum_{k=1}^{M} q_k * \psi(x, y, z; t) \quad \text{(I.8)}$$

Relation I.8 is not very expressive and should be rewritten taking into account that the concentration in a point and in an instant depends on all the particles previously present in the air, emitted in a generic initial instant t0 and able to reach the point (x, y, z) at time t. Basically, these particles have undergone a transition from (x_0, y_0, z_0) at time t_0 to position (x, y, z) at time t and it is possible to define the transition probability Q (x, y, z, t | x_0, y_0, z_0, t_0). Furthermore, it can be considered that the concentration in a point at time t depends on all those particles emitted at the previous instants t', described by an emission probability S (x', y ', z', t ') and able to perform a transition at the point (x, y, z) at time t. If we further hypothesize, for simplicity, that the mass represented by each particle is unitary, the general relationship that describes the average concentration of a pollutant at time t in a point of space is given by a function that integrates the contribution the concentration deriving from all the particles already present at time t_0 and the contribution to the concentration deriving from the particles that are emitted from t_0 to t.

I.2.5 Data interpolation.

The creation of maps of pollutants concentrations to describe the area in a timely manner (the individual cells) continuously show the pollution levels, allowing to appreciate the surface trend of the parameter studied. For maps to be useful, a correct correspondence must be established between the cells and the physical entities they want to describe. A map that wants to represent the trend of a size on the territory is built by assigning to each point the value of the size that belongs to it based on the corresponding geographic coordinates. However, in reality it almost always happens that the values for a limited set of points are known, which corresponds to a more or less small fraction of the totality of the points; and it may they are not uniformly distributed on the image grid. Nevertheless, it is almost always possible to construct a georeferenced map that plausibly reproduces the continuous spatial trend, as there are numerous geographic interpolation methods to populate all the points of the image starting from known ones. Understandably, any interpolating method will be the more effective the more numerous the starting data are and the more uniform they will be

arranged within the points matrix. Spatial interpolation techniques are estimation procedures that make it possible to evaluate the values assumed by a certain quantity on a continuous domain (surface, volume) starting from the known values (calculated or measured) in some points. The result of the spatial interpolation is therefore a surface that represents the values assumed by the size in space. The interpolation can be carried out globally, considering all known values in the domain of interest, or locally, considering the appropriate neighborhoods of the already defined values.

The multi-source Gaussian plume model provides for the implementation of a system that considers each single measured value as a source that is dispersed into the environment according to a Gaussian plume distribution (Gustafson et al., 1977). The multisource plume model is among the models widely used for mathematical modelling pollutants concentrations in urban areas. The Gaussian plume model is the common element of many dispersion models starting from the simplest with a single source up to including many sources. The spatial pollutants in these models is affected by the influence of the parameters from a physical-meteorological point of view. The multi-source model is then developed by superimposing the individual plumes from all the emissions considered.

The latter are then in practice idealized as being large point sources elevating or continuing at different heights above ground level (Calder 1977). Traditional multi-source models consider the integration or sum of all sources. If the source physical properties are considered, such as height and diameter and the variables associated with the emission such as gas velocity, temperature, etc., it is possible to combine all the sources in a single superstack. This superstack is based on the concept of the multi-component system mix. This type of model was used by Ahmad and Bouhamra (1993) to predict SO2 concentrations in an urban and industrial context in Kuwait, providing a better prediction in pollutant concentrations than traditional models (Ahmad and Bouhamra, 1993). New spatialization models developed are based on the pollutant concentration values measured by air quality monitoring systems and tends to reconstruct and locate the pollution source above all in presence of accidental contaminant release (citare Plume e Turbelin et al., 2014).

Among the possible models there are the deterministic methods that consider the link between neighboring points as a function of a law whose parameters have physical significance. These methods include the inverse of distances (Inverce Distance Weighted – IDW), the polynomial interpolations (Spline) and the polygons of influence (Nearest Neighbor Analysis – NNR). In general, these methods do not consider the statistical properties of the measured points, but assign a unique value to the function under consideration, by virtue of the variations of physical parameters, in the study domain. This implies sufficient knowledge of the surface to be modeled.

Another important aspect is that they do not carry out error estimates on the results obtained.

The missing values can be estimated also by using a geostatistical model. The geostatistical methods (kriging) are based on the measurement of the spatial autocorrelation of the values, regardless of the possible physical meaning, and include an evaluation of the forecast error. Geostatistics, in fact, studies the behavior of variables, their self and mutual spatial correlations, their structure, extracts the rules according to coherent models and uses them to carry out the operations that are required to solve specific problems. The methods of Geostatistics are applicable in all fields of applied sciences in which the phenomena of study are characterized by the character of spatiality. The methodology of Kriging is well known in literature but there are very few works that implemented it for large datasets as those treated in this case, due the calculative burden increasing more than linearly with the data set size. Ordinary kriging models are also employed to determine the spatio-temporal variety of annual and seasonal PM2.5 and PM10 concentrations (Karimi and Shokrinezhad, 2021)

Another possibility to estimate air quality in the missing cells, is to use measured data in the neighbouring cells with purposely designed algorithms (Lotrecchiano et al., 2021). In the present study, the results of the applications of few simple methods is assessed by comparing, in each cell where available data are present, the values predicted by the algorithm using the neighbouring data end the experimental value. Different criteria will be adopted and compared. In some of them the influence of the neighbouring cells on the prediction of the local concentration is based exclusively on the distance between the centre of the influencing cell and the centre of the tested cell, in others the direction and the speed of the wind are taken into consideration in the calculation procedure. The approach proposed involves various innovative aspects, in fact this study use air quality data provided by an extremely innovative technology really and currently measured in an Italian city. Moreover, the model is applied to a very large area. Finally, the developed method for the spatial estimation of missing data was compared with other tecniques to evaluate the best solution.

I.2.6 Fires monitoring.

Fires are accidental events determined by different causes, produced by natural events, such as accumulation of gases due to the decomposition of organic matters exposed to high temperatures, lightning, etc. or by human actions for unexpected, legitimate or illicit reasons.

Gases, smoke and vapors that develop during the non-controlled combustion occurring in a fire can also be responsible for significant damages to both humans and the environment. Smoke is perhaps the most visible part of the product of a fire. It can be made of very fine solid particles suspended in the

gas phase, as well as of condensed vapors. Gases produced in a fire are able to maintain their state also when cooling at ambient temperature. Vapors leaving the fire can become solid or liquid at ambient temperature and, when they move away from the flame, may condense and adhere to cold surfaces or form particles (aerosols) that remain suspended and moved by wind. Gases and vapors produced in a fire and coming in contact with humans can cause discomfort to breathe, irritation, interact with skin or internal tissues, such as nasal mucous membranes, lungs, organs, and internal organs, producing dangerous effects due to long exposures (Fasogbon et al., 2019).

Smoke, gases and vapors developed in a fire in an industrial site may be more or less dangerous in relation to the kind of materials involved, especially if they include potential sources of toxic compounds such as it happens for paints and polymeric materials. When such compounds are released in a fire, they can spread in the air to the ground and to the waterways of the areas surrounding the fire. The main gas and vapor compounds produced as a result of a fire which are responsible of pathological effects are: carbon monoxide (CO); carbon dioxide (CO_2); hydrogen cyanide (HCN); hydrochloric acid (HCl); nitrogen oxides (nitrous oxide - N_2O -, nitric - NO -, and two forms of dioxide - NO_2 and N_2O_4); dioxins and furans. For most of these compounds both the amount released in a fire and the spatial extent of the contamination produced can be evaluated by using different models which make use of the composition of the burnt material, the fire conditions and the transport and dispersion in the air. For these compounds the toxic effect, mainly linked to the presence of the cloud released by the fire, often ceases with the dissolution of the cloud. However, dioxins can deposit on the plants, on the soil, in the water courses and can be ingested by the fauna that populates these areas, accumulating in their organisms (Steenland et al., 2004). Such situation has determined the need to study the possible mechanisms of dioxin formation in order to assess the effects of contamination. The type of smoke produced in a fire depends on the type and on the state of aggregation of the fire fuel, of the type of ignition and the mode of combustion that develops in relation to the ventilation conditions of the fire itself. The state of air quality in a town, especially after important events such as fires, is the result of a complex combination of a multitude of physical and chemical transformations. Physical transformations are mostly related to transport processes determined by air motion, which tends to disperse, convey and deposit the chemical species directly emitted by pollution sources, the so called primary pollutants. Chemical-physical transformations, instead, can be responsible of the formation of new polluting species, called secondary pollutants. The dispersion of pollutants, caused by air instability (vertical dispersion) and transport of air masses (horizontal convection and dispersion), as well as their chemical transformation or deposits are strictly dependent on the dynamic behavior of the lower layers of the atmosphere. Therefore, in order

to draw a clear picture of the extent of the local pollution related to a fire and its impact in the surrounding areas, it is important to know both the point of origin of a fire and the environmental parameters which allow to integrate the space and time distribution of the pollutants.

Due to the accidental nature of a fire, environmental monitoring related to a fire cannot be scheduled. Furthermore, given the variety of substances that can be potentially burned in a fire and the different territorial characteristics of the areas that may be affected, it is not possible to define in advance a standardized monitoring protocol to be applied in case of fire. Even if it is always possible to identify some recurring phases of the monitoring activity connected to the evolution of the fire, the measurement campaign will still have to be designed based on the event, the materials and substances affected by the fire, the place and its environmental characteristics and sensitivity. In order to have a complete view of the environmental situation, it is recommended to install in urban areas air quality monitoring networks, which, beside fixed monitoring stations (Sofia et al., 2018), could also include mobile monitoring stations of the new generations, like Real-time on-road monitoring stations (Lotrecchiano et al., 2018). Such kind of network ensures high space-time resolution of air quality measurements, which can be used not only to define the levels of pollution but also to trace back the pollution and define its sources (Sofia et al., 2018), Furthermore, in case of environmental accidents such as fires, the above-mentioned networks allow to instantaneously evaluate the effects of the event on the air quality and help the authorities and the citizens to better understand the actual environmental situation. In case of everyday air monitoring, the immediate knowledge of pollution levels and their origins can lead to the definition of the most appropriate strategies for their reduction, with reference to mitigation strategies that involve the decarbonization of the energy sources. In fact, these latter combine the reduction of environmental pollution with social benefits (Sofia et al., 2019).

Modelling has been often used to evaluate concentration and dispersion process of pollutants in air (Lotrecchiano et al., 2019), water (Pengpom. et al., 2019) and ground (Tshehla et al., 2019), and dispersion models are useful tools to determine long-term health impacts of the smoke exposures of the local populations (Uda et al., 2019). For example, during accidents like fires, the correct modelling of the pollutant distribution in space and time is fundamental for the definition of the human exposure to pollution. In addition to this, the application of dispersion models can have an important role in providing air quality data to promote the social acceptance to reassure of industrial plants with controlled pollutant emissions (Giuliano et al., 2018). As reported by the recent literature survey (Leelőssy et al., 2014), the dispersion of air pollutants in the atmosphere has been modelled by means of three different approaches: Gaussian dispersion models, Lagrangian models and Eulerian models. Recent studies using the Gaussian dispersion

modelling approach were focused on improving the plume rise mechanisms description by means of semi-empirical Equations or by more sophisticate models accounting for cloud microphysics and transport (Daly et al., 2012, Sofiev et al., 2012, Mallia et al., 2018). Models coupling fire plume dynamics and weather forecast were also developed mainly to forecast the fire plume dispersion (Mandel et al.., 2011). Daly et al. (2012) estimated the ground-level exposure in the neighborhood of a fire in a warehouse, by means of two plume dispersion modeling techniques. In particular, in the study the air quality impact of large fires was assessed by assuming the smoke plume as a combination of a hot upper-part staying aloft and a cooler lower-part impacting the ground. Mallia et al. (2018) highlighted that emissions injection schemes corresponding to the assumption of both no plume rise and plume-top configurations could not predict satisfactorily the vertical distribution of smoke.

The Lagrangian particle model approach was used by several recent studies (Zhu et al., 2018, Ferrero et al., 2019, Thakur et al., 2019). Ferrero et al. (2019) modelled different schemes of plume rise of smoke emitted from controlled burns without any a priori assumption on the dynamics of the plume or on the atmospheric conditions. Good agreement was obtained by Thakur et al. (2019) between model predictions of vertical profiles of CO concentration and satellite observations over the burned areas. Results highlighted a significant increase of CO concentration at altitudes about 3 km height above mean sea level during the fire event. Zhu et al. (2018) combined a Lagrangian Flexible Particle dispersion model (FLEXPART) (Stohl et al., 1998) with the Weather Research and Forecasting (WRF) model output (Michalekes et al., 2001) to derive the smoke aerosols transported from forest fires. Some recent works on fires (Price et al., 2016, Luhar et al., 2020) used the TAPM model (Hurley et al., 2005) based on fundamental Equations of atmospheric flow, thermodynamics, moisture conservation, turbulence and dispersion, combining different sub module approaches (integrated plume rise, Lagrangian particle, building wake, and Eulerian grid modules). A chemical transport model was also combined with a Dispersive Apportionment of Source Impacts (DASI) by Huang et al. (2020) (Huang et al., 2020) to forecast the air quality impact of prescribed fires.

The availability of a model validated with direct experimental measurements of pollutants concentrations is of paramount importance to predict the impact on air quality in the surrounding of prolonged fires and to implement precautionary measures for the population. However, few published studies have validated model predictions with direct measurement of the spatial distribution of pollutant concentration over the time (Luhar et al., 2008).

I.2.7 Transnational migration of particulates.

Particulate matter of anthropogenic origin largely belongs to the fraction with a diameter less than 2.5 μm, while desert clouds of dust contain a significant fraction with larger particle size; a considerable decrease in the PM2.5/PM10 ratio is, therefore, a contributing index of natural particulate matter of desert origin (Steenland et al., 2004). Saharan episodes are characterized by a large increase in the PM10 concentration, not followed by that of PM2.5 and PM1, with a consequent decrease in the PM2.5/PM10 ratio, an increase of the crustal elements (Al, Si, Ca, Ti, Fe) in the PM10 fraction, and an increase in the Si/Fe, Al/Fe, and Ti/Fe ratios (Sofia et al., 2018a, Lotrecchiano et al., 2018). Moreover, the effects of these Saharan events have repercussions, not only for outdoor air quality, but also for indoor air quality, which is affected by the high concentrations of aero-dispersed dust (Sofia et al., 2018b).

From a climatic point of view, the Mediterranean atmosphere (MED), as defined by Jeftic et al. (1992), is characterized by rainy winter seasons and hot and dry summer seasons, which affect the continental areas that surround it. It is therefore full of particulates of anthropogenic origin, coming from the vast industrialized European regions, and of natural particulates of crustal origin, coming from the extensive arid and semi-arid areas of Africa and the Middle East, and of aerosols of marine origin, generated by the surface of the MED, and of volcanic dust, emitted by the main volcanoes of the basin, such as Etna and Stromboli (Arnold et al., 1982). From the studies carried out on the influence of Saharan dust on a large scale, it is highlighted that latitude is a very important factor that must be taken into account in studies concerning PM10 concentrations. This same difference can be seen between the number of Saharan transport events that take place in Spain and those that take place in Italy or England (Ryall et al., 2000). The possibility of the occurring an interference of the powder transport in the PM10 concentration is higher in Spain than in northern Italy, and much greater than in England. Various studies have involved Italy in recent years, highlighting how, depending on the areas investigated, the days of transport of dust increase from north to south (Chiari et al., 2004). Therefore, a need is highlighted, concerning European countries located at higher latitudes, to look for these signals, both in PM10 and PM2.5. The problem of dust can be analyzed using integrated approaches to define the link between the natural and the urban dust within the environment (Doronzo and Dousari, 2019). Unlike the mineral aerosol which is of natural origin, the PM10 produced in the urban environment comes from different anthropogenic sources. The major particulate matter sources in urban and extra-urban areas are represented by industrial plants, domestic heating systems, emissions due to vehicular traffic, tire breaking, agriculture, and livestock.

I.2.8 Artificial intelligence.

For this to happen, the machine set up for an artificial intelligence (AI) must be trained. Training takes place by showing the machine thousands data also representing the subject of interest, and labeling them by category. In this way, when the machine is shown a new data, then it will return the label of the data that have instructed the machine with more characteristics coinciding with the one to be recognized. This type of training is called supervised machine learning as you already know the category to which the images instructing the machine belong. The means used by learning systems are neural networks. Through external data, they can change their structure by modifying the links between the nodes, thus making the system adaptive. The learning and reasoning phase is therefore given by the connection of internal information and external data, through the neural network. The most used neural networks in image recognition are convolutional neural networks (CNN or ConvNets) and are precisely those based on the behavior described above. In the field of machine learning, a feature is an individual property and is measurable for an observed phenomenon (Bishop, 2006). It is the result of a processing aimed at transforming the information of a pattern into information of a different type, more precise and therefore more suitable for carrying out the pattern recognition task. This processing is called dimensionality reduction and is divided into feature extraction and feature selection. Feature learning is defined as the set of techniques that allow a system to independently discover which features to extract from the input patterns. It turns out to be one of the main innovations of machine learning, replacing learning engineering, in which it was man who chose the common characteristics to be used. Feature learning is divided into different types: supervised learning, unsupervised learning, semi-supervised learning and reinforcement learning.

Supervised learning is a machine learning technique that aims to train a computer so that it can perform tasks automatically. In this type of learning, examples are given to the system as input and the correct results that the system should aim for. The system will gradually learn to adapt its features, so that if we apply a pattern at the input, the output will be what is expected (clearly with a percentage of error). Such learning is used, in most cases, in classifications and regressions. Generally, an AI system transforms the input data into output data through an objective function F, regulated by a set of parameters θ Learning consists in discovering the optimal parameters θ'for which, by applying a function F to a set of input values, an output data equal to those expected, net of a certain error ε is obtained. A dataset is a collection of data, which can be of various kinds. A training set is a set of data to be provided to the model in order to train it and optimize the parameters θ. A test set is a set of data to be provided to the model to evaluate the performance of the chosen hyperparameters. A validation set is

a set of data, with which we evaluate the final performance of the model and the degree of generalization achieved. Obviously, the three sets must be separated from each other in order to avoid overestimation of performance and overfitting. This is because, if the validation or test set are subcategories of the training set, it means that the examples provided in the evaluation phase have already been seen and learned by the model which will therefore provide good results. Hyperparameters are values that characterize the learning of our model. These hyperparameters will affect the functioning and performance of our model. The hyperparameters that are commonly studied are Epochs, Batch Size and Learning Rate. An epoch refers to a cycle through the entire training dataset, which also means that every example in the training dataset has had a chance to update internal model parameters. The batch size indicates the number of samples the system works on before updating internal parameters. A simplified view would be to see epochs and batch size as two nested for loops, the outermost iterating over the epochs and the innermost over the batch size.
The Learning Rate is a hyperparameter that indicates how fast or how slowly our model is learning. After optimizing the learning rate, our model will learn to approximate, in the best way, the function, given the available resources, in a number of training epochs. The learning rate can be considered the most important of the hyperparameters. To evaluate the results achieved in an AI system, the distance between the expected result and the result that the system outputs is measured. To measure this distance and therefore to evaluate the response of the system, we use metrics, which are tools of a mathematical and/or statistical nature. As regards the classification, reference is made to the Loss Function and Accuracy. The loss function is a function that evaluates the performance of a model by describing the loss of information. To be considered performing, a model must minimize this function: in general, the smaller the loss, the more performing the model will be (except if the model overfits) and is calculated on the training set and on the test set. As for accuracy, we can define it as the complement of the loss. This function is generated on all three sets and the percentage of patterns that can correctly recognize the total ones is calculated. The confusion matrix is a matrix showing the true classes on the rows and the predicted classes on the columns. It is a great tool for classification, as it allows you to see in detail where the classification errors are distributed. A good confusion matrix should report high values on the main diagonal and low values in all other cells.
Looking at the definitions of loss function and accuracy, it can be understood that a good classification is considered as such if both approximately converge. At each epoch, the accuracy will have to converge to approximately one, while the function loss will have to converge to approximately zero.

Overfitting occurs when in the training phase the model learns very well and is unable to generalize on the test and validation phases. The model fails to learn the general pattern recognition rules. This can happen for several reasons, such as: training set too small or not very varied, the model is too complex etc. If the model goes underfitting, it fails to learn well in the training phase and cannot even generalize new data in the other two phases. A model that goes into underfitting is a model to be discarded, it can be seen already in the training phase, therefore it is often not even discussed, due to its simplicity of detection.

The generalization capacity of a model is the ability of a model to transfer the high accuracy achieved in the training phase to the test phase. Therefore, the generalization refers to the way in which the concepts learned from a model, during learning, are applied to specific examples not seen by the model during the training phase.

I.2.9 Thesis objectives and structure.

The literature available on the environmental field is based on few hundreds of data and all the analysis are made with a macroscopic approach. This work aims to valorize the thousands of data provided by the high time-space resolution networks implemented by Sense Square integrated with other data sources as the Regional Agency for the Environmental Protection (ARPA) and Copernicus Sentinel 5 and, use it to characterize the Italian environmental matrix.

The thesis aims to analyze data available from different measurements technologies from a modelistic point of view to understand the behavior of the aero-dispersed pollutants and used to obtain information on the air quality status.

The data from the fixed monitoring stations will be used for the optimization of the sampling points number and positions which is the first crucial thing in the network design. Also they can be used for the pollutant tracking in fact, matching the high quantity of data available from each sensing device with the meteorological parameters measured together with the particular mesh in which the monitoring stations are located allows the knowledge not only of the pollution levels but also the individuation of the pollution sources. Also the data could be used to study the seasonality variability of pollution both in urban and extra-urban context. Part of the thesis work aims at using a modelling approach to derive the dispersion of pollutants emitted by an accidental fire validating the results with measured data regarding air pollution. The Gaussian plume dispersion approach is followed in order to evaluate the mass flow rate of particulate matter originating from a fire by using a limited number of PM2.5 concentration values measured by an air quality monitoring station placed 4 km downwind of the fire starting point.

Moreover, it will be possible using multiple linear regressions or some Bayesian or machine learning approach, to train a model with the data measured and build a forecast model for the pollution.

Data have been also used to investigate the real situation of air quality during the first lockdown due to Sars-Cov2 pandemic even and to evaluate the contribution of Saharan dust to the environmental
pollution measured by the air quality monitoring stations in an urban context.

The real-time on road monitoring stations are a very innovative way to measure the air pollution. The literature available for this new technology is poor in the data analysis. In fact, the dynamic network described in literature are composed of few units while in the network studied in this work the sensing units are 53. Moreover, the sensing device used in this case have a 3 minutes' resolution so, the data provided by the network are enormous. All this data allows the detailed knowledge of the environmental situation with a high space resolution. The availability of data from the dynamic network is strictly connected with the vans operating hours from 09:00 to 17:00. Moreover, the urban context is full of traffic restrictions like traffic limited zones, pedestrian pathways and parks that doesn't allows the vans passage. So, it is possible that some of the cells in the grid which the city is divided into have no data available. To overcome this lack of information, it is recommended to install also a fixed monitoring network. With the hybrid network it is possible to integrate the data coming from the two types of networks to have an overall view of the environmental situation. To fill the gap of the missing values it will be possible to implement some techniques to reconstruct evaluate the missing values. This particular techniques, includes geo-spatialization methods and multiple imputation of the missing values. The geo-spatialization methods are well known and widely used in literature by are not applied to large dataset as the one used for this work. The core of the work consists in the development of a model for the air quality data spatial interpolation described in the Chapter IV.6.2 and IV.7.

The different characteristics of the data coming from the fixed and dynamic monitoring networks require a different approach to the data. A deterministic approach has been applied predominantly to the data of the fixed networks while a stochastic modelling has been applied to the dynamic ones. The monitoring campaigns involving the fixed monitoring networks have been used to evaluate the dispersion model and to define the pollutant temporal provisional algorithm in terms of dispersion. Data from the dynamic networks have been used to develop a model for the spatial provisions of pollutants evaluating the missing data in the space.

All the methodologies used and the results obtained in the different data analysis are reported in Chapter IV and each section has been published on an international research journal.

Chapter I
Experimental apparatus

I.1 Fixed monitoring stations

Each monitoring network is composed of air quality sensors based on IoT technology. The sensors placed inside the control unit (Fig. I.1) are able to monitor PM10, PM2.5 and PM1 particulate, NO_2, O_3, CO, H_2S, SO_2 and the Organic Volatile Compounds (VOC). To complete the measurement unit there is a weather station which measures temperature, pressure, relative humidity, wind intensity and direction. The data recorded every three minutes are sent in real time to a database and available online and on a dedicated smartphone application.

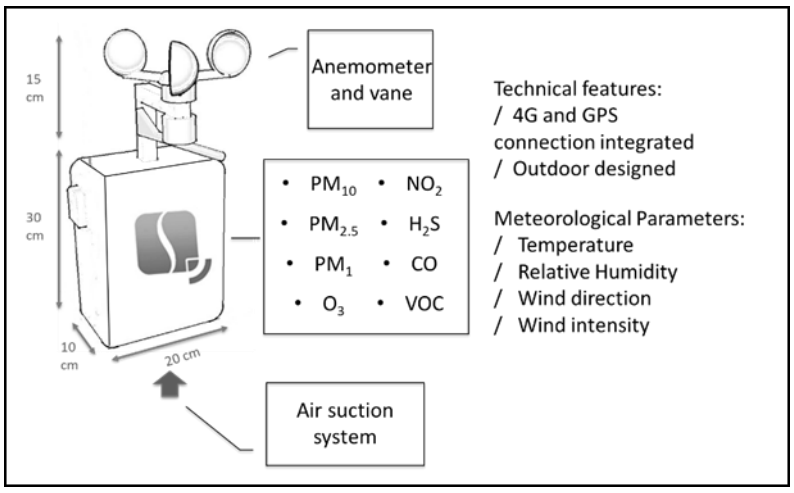


Figure I.1. *Measuring device.*

The sensors used are different depending on the pollutant analysed. The one used to measure particulates concentrations is based on the principle of laser scattering. In these sensors, the airborne particles, hit by the laser beam, diffract light radiation in different directions depending on the particle size. Therefore, detailed information on the concentration of the different particle size of the powder can be obtained. In particular, the built in software of the sensor elaborates the signal in order to provide cumulative concentrations of particles smaller than 10 µm (PM10), and smaller than 2.5 µm (PM2.5). The sensor chosen to detect NO_2 and the other gases concentration is based on electrochemical principles (Kuberský et al., 2014). The measurement range is between 0 and 200 ppb (volume) with a resolution lower than 20 ppb and a measurement repeatability within ± 3 % of the full scale. The VOC sensor, based on tin oxide sensing component, measures the concentration of the mixed equivalent organic compounds (Mirzaei et al., 2016). The basic solution of the measuring device designed and patent by Sense Square can be easily modified to be implemented in a fixed or dynamic network. One of the most important things during the design of the monitoring network is the definition of the correct locations of the sampling points where to install the sensors. In this perspective, an algorithm for the definition of the optimal sampling positions has been implemented using a modellistic approach. The sensor characteristics are reported in Table I.1

Table I.1. *Sensor characteristics.*

Topic	Parameters	Sensitivity	Technology
Climate	Temperature	±0.3 °C	Band-Gap Sensor
	Relative Humidity	±2 %	Capacitive Sensor
Particulate	PM10	±5 µg/m^3	Laser Scattering
	PM2.5	±5 µg/m^3	Laser Scattering
	PM1	±5 µg/m^3	Laser Scattering
Gas	CO	--- µg/m^3	MOx
	NOx	--- µg/m^3	MOx
Wind	Intensity	±5% m/s	Ultrasonic
	Direction	±0.3 °	Ultrasonic
Radiation	Radiation	- W/m^2	Pirometer

Humidity can influence the particulate concentrations measurement, altering the result. The laser could in fact confuse the water particle with the dust particle and thus alter the measurement of the particulate concentration. To overcome this problem, the air entering the sensor, before entering the measurement chamber, is heated to eliminate the humidity inside it to have a reliable measurement. Over time, the growing development of sensor technologies has led to sensors with high sensitivity, capable of making very precise measurements. From table I.1 it can be seen that all the sensors used have very high sensitivities. The technical difficulty in integrating each sensor into a measurement device that can provide information on several

types of pollutants has been solved by carrying out an important design study and engineering phase of the instrument development and following the sensors firmware implementation for each type of sensor. Furthermore, the installation in sensitive points of the cities is facilitated by the compactness of the measurement device. Obviously it requires the use of an internet connection system which can be integrated internally or be shared with the external environment via a LAN cable or wireless connection. Furthermore, measurement device requires a power supply which is commonly represented by a domestic power socket but which in some cases can be represented by batteries or integrated photovoltaic panels. The system has been carefully designed and patented (WIPO patent 2018 / 225030A1).

I.2 Real-time on-road monitoring stations

The mobile monitoring unit was built so that the suction part (Fig. I.2b) is located in the upper part of the rear of the van. This position allows to minimize the direct influence on the air sampling of the air flow produced by the vehicle movement. In the on-road version, the sensor has been modified to ensure that the sampling rate is isokinetic to that of the vehicle to avoid competitive motions between the phenomena. Furthermore, the positioning of the air suction area in the upper part of the vehicle, obtained after multiple field tests, prevents the pollutant measurements from being directly influenced by the vehicle's own emissions which could alter the experimental measurements.

In order to measure concentrations of particulates, NO_2 and Organic Volatile Compounds (VOC), the mobile stations are equipped with three sensors. Each of the three sensor is fed by a dedicated air stream generated from the incoming airflow by an air spitting element.

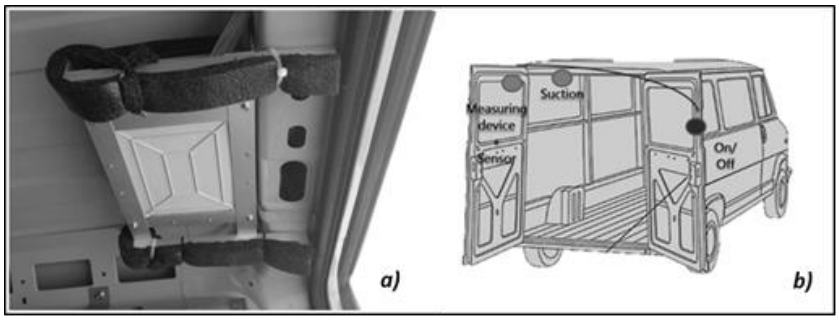


Figure I.2. *a) Measuring device lodged on the van. b) Schematic representation of the measuring system.*

All mobile monitoring stations are equipped with a mobile meteorological station, which takes records of temperature, relative humidity, and pressure. A GPS unit provides the information on the position

of the sensing unit in terms of latitude and longitude. Aggregated data of the sensed concentrations, meteorological information and position are sent to the central server using IoT technologies. The data are collected and stored in real time (Castell et al., 2018) continuously for the whole working time of the van every 3 minutes. The on-road mobile network is composed of 53 fully working stations that every day pass along Milan each day and collect data.

Vans belong to the urban network of a mail and parcel delivering company. According to the needs of an effective delivering action, each van has a well-defined route in a limited area.

For each day around 63600 data are collected by the network and this data have to be aggregated in the way to have a clear view, to be easily analysed and, to be explanatory of the environmental matrix. In order to aggregate data two different strategies were adopted. In one case large areas for data aggregation, surely including a statistically relevant number of data were considered. In the other cases a finer ground subdivision was adopted in order to reduce the granularity of the data. In this latter case the significance of the data associated to each of the ground subdivision will be an object of investigation. The large aggregation areas for the data were coincident with the nine administrative districts of Milan (Fig. I.3).

Figure I.3. *Milan districts.*

Numbering of districts, starts from the city center and increment in the clockwise direction in the peripheral districts. A preliminary analysis of the distribution of the vehicles on the ground and of the routes covered by vans was performed in order to identify the most significant areas for the environmental survey. The data analysis with finer spatial resolution was obtained by dividing the city map into square cells of 1 km sides represented

as parallelepiped on the map reported in Fig.I.4 due to the distortion of the Mercatore projection used. The centroids of each cell form a regular grid, and the cell area of 1 km^2 turned out to be the minimum area to aggregate a sufficient number of data, to accurately describe the complex urban context. The km^2 aggregation is the right compromise between the high time and space resolution. In fact, the high time resolution provides a lot of data but if they are not correctly limited, they could result too dispersed and not useful for the analysis. Moreover, the dimension of km^2 should contain around five streets that is the right compromise between an area street crowded and a poor street area.

Figure I.4. *Grid map made of square cell of 1 km side used to describe the pollutant concentration distribution on the ground.*

Chapter II
Materials and methods

II.1 Matlab data management

Sense Square's high-resolution networks consisting of a minimum of 3 control units that provide approximately 15,000 data for each control unit, making it necessary to implement a data manager system. The instrument chosen for data management was Matlab, which is able to communicate directly with the ThingSpeak IoT server on which data are sent.

The implemented system allows the quick definition of the hourly and monthly trends of all the parameters measured by the control units such as PM1, PM2.5, PM10, VOC, NO2, temperature, pressure, relative humidity, wind direction and intensity. The measured parameters are all scalar variables except for the wind direction which is a circular variable and as such is appropriately defined in the average calculation.
Hourly and monthly trends are plotted on charts that have been implemented within the system, highlighting any exceedances with respect to the law limits imposed by D. Lgs.155/2010.

Monthly reports are produced for each monitoring network, which contain the graph of the aforementioned parameters and comments on pollution trends. Furthermore, the combination of pollutant concentration data and weather parameters allows the definition of the pollution direction so that it can be more detailed and define not only the level of pollution but also its source.

II.2 ROMs network data management

The on-road monitoring network requires the implementation of a different data management system. In this case the data are sent directly to the Firebase database from which they are periodically downloaded and processed in Matlab. The admin interface implemented in Python allows the

direct download of the data directly from the database. In this case the excel or csv file resulting has a matrix containing the timestamp, the name of the district which the data belongs to, name of the km^2 considered, PM1, PM10, PM2.5, NO_2, VOC concentrations, traffic intensity, temperature, pressure, relative humidity, latitude, longitude and number of sensor which the data came from. The data can be downloaded as raw data or aggregated in hourly and daily averages. Usually the data used for the basic analysis are downloaded as raw data and subsequently aggregated via Matlab. The Matlab code allows the definition of the pollutant daily average concentrations starting from the hourly average, obtained from the raw data for each km^2. Also the number of data are counted to verify how many data there are for each cell. For each municipality in which the investigated area is divided, monthly trends are analyzed and, depending on the case, for the municipality of interest a more in-depth analysis is carried out. Analyzing the daily trends, it is possible to highlight first of all possible exceedances compared to the law limit and then it is evaluated the level of exposure of citizens to a high level of pollution. Within the municipality of interest considered, the cells (squares) that are most polluted are identified. For each of these cells even more in-depth analyses are carried out considering the daily trends, depending on the time of the day and according to the day of the week

The GPS integrated in the control units provides position data in terms of latitude and longitude which allows to trace the daily routes of the vans. These routes are extracted in a geojson file that can be viewed online directly on the map of Milan. A report indicates the vans that did not circulate on that day either due to maintenance or to a scheduled stop. The analysis of the routes allows their optimization both by modifying any vans routes that pass outside the area of interest, and by remodulating overlapping vans to move them to areas where coverage is less or even absent.

A further improvement to the reporting system will be carried out by implementing an automatic report generation system which will allow a considerable increase in the efficiency of the management system.

Matlab is also used to implement the Gaussian Plume dispersion models. In particular, the GaussianPlume function allows the calculation of the spatial pollutant concentration given a certain source setting the custom parameters.

Chapter III
Modelling and results

III.1. Prediction of pollution transport with OSPM approach

Model

The Operational Street Pollution Model (OSPM) was used to evaluate the PM10 concentration on the ground. The OSPM model, developed by Hertel and Berkowicz (1989), is a model of representation of pollutants based on the description of their dispersion and flows in street canyons. The Equations used by the model to evaluate the PM10 concentration rely on various parameters that are specific of each street.
Briefly, the main components of OSPM are summarized as follows. For the pollutant PM10, the concentrations are modelled as:

$$C = C_{bg} + Q_{trf}\ (C^*_{dir} + C^*_{rec}) \qquad \text{(III.1.1)}$$

where:

- C is the concentration at the receptor point (μg m-3)
- Cbg is the background concentration (μg m-3)
- Qtrf are the emissions from traffic per street length (μg m-1 s-1)
- C* is a function determining the persistence of pollutant including the effect of meteorology, traffic speed and volume and street geometry (units: s m-2). It is composed of a direct, C*dir, and a recirculation, C^*_{rec}, component.

The concentration of PM10 mainly includes the pollutants emitted by vehicles and combustion processes (industrial or house-heating combustion emissions). In the C* function the C^*_{dir} term is calculated assuming a Gaussian plume which represents an external source, while a box model is used to calculate the contribution from the recirculation vortex (C^*_{rec}). In particular, the recirculation vortexes are formed only under certain conditions, governed by the wind speed and direction, as well as the canyon

geometry. Meteorological parameters as the wind speed and direction are used to determine the size of the recirculation zone. On the windward side, C* includes only the recirculating component (possibly with a contribution from the direct component, depending on the length of the recirculation zone) while on the leeward side C* is the sum of the direct and recirculating components. Both sides account for the urban background. The function C* includes also the contribution from the turbulence produced by the vehicular traffic, which depends on the speed and the intensity the
traffic, as well as from the frontal area of vehicles. Meandering of the wind direction is also accounted for in C*, as the wind direction is not expected to remain constant during the 24 h time steps.

Background input

For all of the modelled pollutants, the OSPM model requires a background value in input. The value of the urban background can be obtained in several ways such as using a simple large scale dispersion model to evaluate it or measuring the concentration of the pollutant. In this work, the urban background for the PM10 concentrations was evaluated using a monitoring station of the Regional Agency for the Environmental Protection (ARPAC). The ARPAC monitoring station is named as an urban-background station. In fact, it is located in a point free from the influence of direct sources of local pollution. In this way the only type of pollution considered is carried by long distance transport mechanism and this value represents the fraction of pollutant not dependent from the modelled dispersion.

Traffic input and Fuel input

An important parameter to evaluate the pollutant concentration and dispersion with the OSPM model is the traffic and other related information. The number of vehicles of the different classes (passenger cars, vans, motorcycles, buses, commercial trucks), the mean vehicular velocity and the hourly traffic volumes were provided to the model for the period of investigation to evaluate the influence of the traffic. As an example, the values of the numbers of vehicles divided by type crossing one of the considered streets in the hour interval is reported in Table III.1.1. In particular, the data for the traffic in Battipaglia were exported from the national vehicle database, which accounts also for the emission class of the vehicle, and used for the concentration evaluation. The vehicular emission factors were calculated by using the OSPM tool for the emissions (Tab. III.1.2).

Table III.1.1. *Number of vehicles divided by type crossing one of the considered street in each hour interval.*

Time interval	Passenger cars	Vans	Truck <32 ton	Truck >32 ton	Buses
00:00-01:00	175	12	2	5	0,48
01:00-02:00	70	7	2	2	0,17
02:00-03:00	43	4	5	1	0,17
03:00-04:00	39	5	3	0,5	0,08
04:00-05:00	83	13	5	10	0,17
05:00-06:00	315	55	12	8	1,04
06:00-07:00	1177	240	37	20	2,30
07:00-08:00	2194	284	62	30	4,16
08:00-09:00	1986	242	71	31	5,10
09:00-10:00	1601	265	68	33,	4,47
10:00-11:00	1561	267	74	31	3,62
11:00-12:00	1697	284	82	26	4,00
12:00-13:00	1745	279	75	29	3,51
13:00-14:00	1840	256	88	35	3,60
14:00-15:00	2156	270	71	32	5,12
15:00-16:00	2566	302	57	28	6,03
16:00-17:00	2905	245	52	21	5,40
17:00-18:00	2322	163	33	18	5,52
18:00-19:00	1723	134,	16	12	3,48
19:00-20:00	1195	104	10	5	2,10
20:00-21:00	854	89	8	5	1,61
21:00-22:00	909	73	6	3	1,44
22:00-23:00	846	71	3	3	0,93
23:00-24:00	469	45	2	5	0,45

Table III.1.2. *Percentage of vehicles divided by fuel type and vehicles.*

	Gasoline	Diesel	LPG
Passenger Cars	81%	17%	2%
Vans	15%	84.9%	0.1%
Trucks <32t	0.79%	99.21%	
Trucks <32t	0.79%	99.21%	
Buses		100%	

The methodology for calculations of emission factors that is used in WinOSPM is based on the European Emission Model COPERT. In the

COPERT model the emission factors are specified for a number of vehicle sub-categories covering the majority of European vehicle types. The emission factors specified for each of the vehicle categories are assumed to be universal, i.e. they don't depend on the Country settings. However, the subdivision into the sub-categories is country specific and depends on the composition of the national fleet of vehicles. Emission coming from vehicles considered, for each type of fuel, the coexistence of different European Emission Standards, in particular for Gasoline and LPG were considered Euro I, Euro II, Euro III and, Euro IV standards, while for Diesel was also considered Euro V standard.

Meteorological data input

The month chosen for the simulation was August and in order to take into account, the influence of the ambient conditions of the points modelled, values of temperature, relative humidity, wind speed, and intensity were used in the development of the OSPM model. As an example, all the data used as input for one of the streets are reported in Table III.3. All the meteorological parameters were taken from the measuring stations of an Italian Meteorological Association located near the background station.

Table III.1.3. *Meteorological data input for one of the streets modelled.*

Day	Temperature [°C]	Pressure [kPa]	Relative humidity [%]	Wind intensity [m/s]	Wind direction
01 August	30	1006	52,27	1,55	ENE
02 August	28	1007	58,05	2,98	NE
03 August	28	1006	60,07	2,60	NNE
04 August	30	1005	59,37	5,87	NE
05 August	30	1006	54,17	4,06	NE
06 August	30	1007	53,41	2,75	ENE
07 August	30	1007	54,32	2,05	E
08 August	30	1007	52,85	2,05	ENE
09 August	30	1008	49,26	1,85	E
10 August	31	1010	55,37	1,92	E
11 August	30	1010	56,38	1,92	E
12 August	30	1009	55,93	1,81	E
13 August	30	1007	52,66	1,80	E
14 August	28	1005	54,57	2,41	ENE
15 August	24	1004	69,98	1,42	NE
16 August	28	1005	63,69	5,47	NE

17 August	28	1006	59,14	3,52	ENE
18 August	28	1007	54,42	2,85	ENE
19 August	29	1007	50,21	4,94	ENE
20 August	27	1005	62,94	7,72	NNE
21 August	28	1005	65,58	7,32	NNE
22 August	28	1006	64,84	7,06	NE
23 August	28	1004	54,37	3,66	NE
24 August	27	1002	56,90	3,27	ENE
25 August	28	1002	55,45	1,92	E
26 August	26	1001	60,52	2,07	SE
27 August	25	1003	46,28	5,44	N
28 August	27	1007	39,16	3,24	NE
29 August	27	1007	37,49	1,55	NE
30 August	27	1008	42,68	0,84	ENE
31 August	27	1009	51,81	1,27	E

Streets description

The city of Battipaglia, the focus of this study, is a town of south-western Italy with a population of almost 50,000 people. The city is known for being one of the most important centers of production of typical agrifood products, which makes it one of the highest intensity production sites of its region. In particular, the city is also characterized by the presence of a major industrial pole, not only related to the transformation of agricultural productions. The sampling point chosen were located in the most representative points of the city. Three points (B2, B3, B4) are located nearby the city center, characterized by a high anthropic activity such as commercial activities and schools. The other streets (B1, B5) are representative of another part of the city as the industrial area and a residential area, commonly found in all cities. All the streets characteristics and geometry are reported respectively in Table III.1.4 and in Figure III.1.1 while the points location on the city map are shown in Figure III.1.2.

Table III.1.4. *Characteristics of the street canyons in the model*.

Street name	Length [m]	Width [m]	Height [m]	Receptor height [m]	Type (respect to receptor)	Zone type
Viale Danimarca (B1)	100	10	10	6	Symmetric	Industrial development area
Via Parmenide (B2)	135	10	10	5	Asymmetric	Commercial area
Via Jemma (B3)	150	9	12	10	Symmetric	City centre
Via Baratta (B4)	100	15	20	10	Symmetric	Outskirt
Via Gonzaga (B5)	97	13	12	6	Asymmetric	Residential area

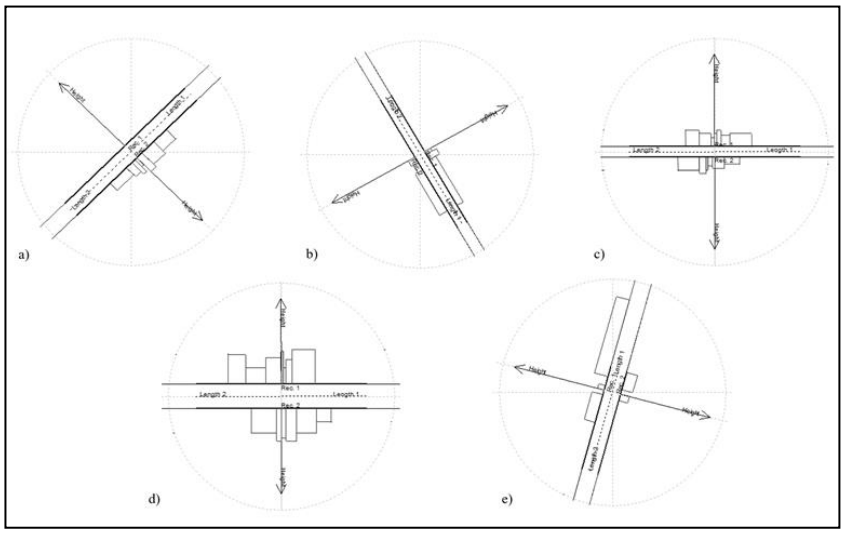


Figure III.1.1. *Geometry of the street canyons in the model: a) Viale Danimarca (B1): 45°; b) Via Parmenide (B2): 150°; c) Via Jemma (B3): 90°; d) Via Baratta (B4): 90°; e) Via Gonzaga (B5): 15°.*

Figure III.1.2. *Air quality monitoring station network located in Battipaglia.*

Results

The output of the OSPM model is given in terms of PM10 concentrations. In order to determine the correct position to install in the future the measuring stations that are representative of the whole city, the concentration profile of PM10 obtained for each of the five positions (B1-B5) selected after the modelling are reported in Figure III.1.3.

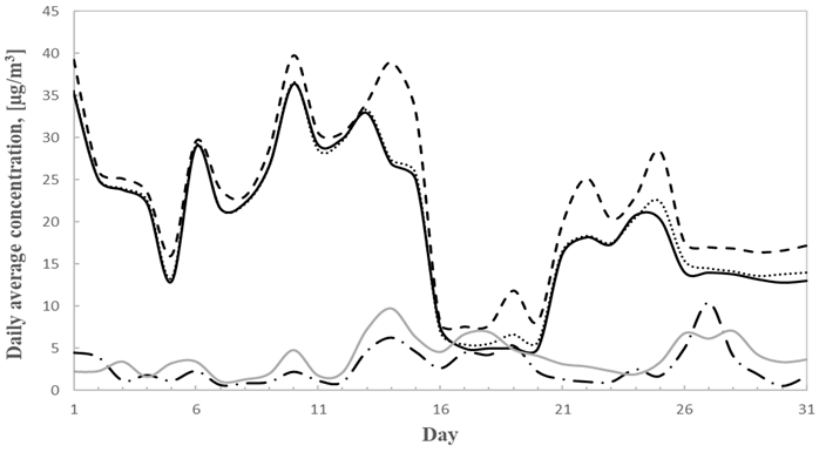


Figure III.1.3. *Daily average concentration of PM10 modelled. The solid black line for B1, the black dashed line is for B2, the black dotted line for B3, the black dash-dot line is for B4, the solid grey line is for B5.*

Figure III.1.3 shows two main features of the results. In particular, the results for the positions B1, B2, B3 show a significantly higher pollution than those of the sampling positions B4 and B5. This difference makes the positions B1, B2, and B3 more interesting in view of a possible pollution sampling station. In general, the more polluted streets are those characterized by a high anthropic activity. The presence of intense traffic, industrial activities and the high number of exposed subjects who attend the activities presents as shops and local market make the positions B1, B2 and B3 the most polluted and the corresponding streets an interesting area to study. Obviously, the theoretical point of view has to be matched with the practical installation of the monitoring stations in fact, not all the interesting points indicated by the model would be easily reachable for the installation. Weak oscillation in the results highlight the weekly periodicity of some of the anthropic activities, perhaps related to traffic changes in the weekdays.

After the installation of the sensing devices in the points individuated by the optimization procedure, the measured values were compared with the modelled ones. The modelled values of PM10 (Fig. III.1.4) are comparable respect to the measured values. The great accordance means that the values predicted by the model were reliable and the modellistic hypothesis were correct.

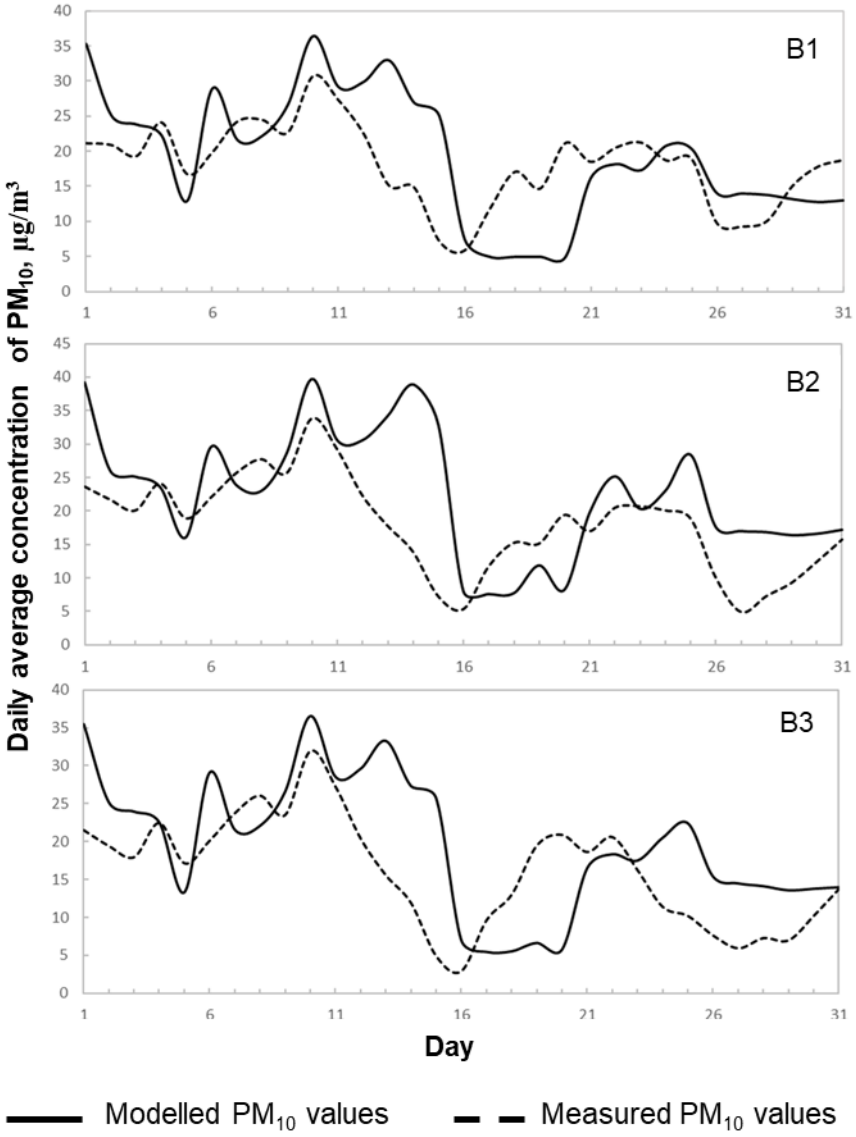


Figure III.1.4. *Daily average concentration of PM10 modelled. The solid black line for B1, the black dashed line is for B2, the black dotted line for B3, the black dash-dot line is for B4, the solid grey line is for B5.*

III.2. An innovative approach to determining the contribution of Saharan Dust to pollution

Air quality data

The control of the air quality parameters according to the regulatory provisions of Legislative Decree 155/2010 represents one of the main institutional activities of the Regional Agency for Environmental Protection of Campania region (ARPAC). The agency manages the monitoring network determined according to regional specifications for the identification and management of air quality monitoring stations, to assess the air quality and polluting emissions spread over the area. The configuration of the network in the Campania region (a southern Italian region) includes 36 fixed monitoring stations and 5 mobile laboratories (Figure III.2.1).

Figure III.2.1. *ARPA air quality monitoring network located in the Campania region.*

The monitoring stations are located in sensitive areas, following the classification of the regional territory. There are also 10 additional fixed monitoring stations installed near the waste treatment plants ("STIR" network) that, although not part of the regional network, provide additional and support measures for the interpretation of the evolutionary phenomena of air quality on a regional basis. The technology on which the measuring instruments of the ARPAC network are based comply with the technical standard UNI EN 12341: 2014. The measurement principle is gravimetry, and the reference method for the determination of PM10 particulate material is based on the collection of the PM10 fraction on a special filter and the

subsequent determination of its mass by gravimetric methods in the laboratory, after conditioning the filter in controlled conditions of temperature (20 °C ± 1) and humidity (50 ± 5%). In Avellino, the ARPA air quality monitoring network is composed of two monitoring stations, V° Circolo (VC) and Scuola Alighieri (SA), classified as suburban background and urban traffic, respectively. Both monitoring stations provide for PM10 and PM2.5 particulate matter measurements with a daily resolution using gravimetric devices and according to the legal specifications. The VC station also provides hourly particulate matter analysis by using laser scatter technology. Both stations also provide NO_2 measurements, and SA adds measurements for CO and benzene, while VC adds only O_3 data.

The main characteristics of the stations considered in this work are summarized in Table III.2.1. The SA station is located on one of the most congested roads of the city, while the VC station is located near a school that represents a sensitive point (see Figure III.2.2). For this work, data for February 2021 were investigated.

Figure III.2.2. *Map of measuring stations located in Avellino.*

Table III.2.1. *Characteristics of the monitoring stations*.

Station	Acronym	Type	Measurements Technique	
V° Circolo	VC	Suburban background	Gravimetric Laser scattering	ARPAC
Scuola Alighieri	SA	Urban traffic	Gravimetric	ARPAC
Mt. Vergine Observatory	MV		Laser scattering	MVOBSV

In the city considered there is also an extra air quality monitoring station, part of the meteorological network managed by the non-profit organization MVOBSV. The sensor used is a PMSA003 Digital Laser Dust Sensor (MV) placed on a palace rooftop (18 m) in the city center (see Figure III.2.2). The sensor is based on laser scattering technology with a time resolution of 10 min. The sensor uses the principle of laser dispersion, that is, by irradiating the suspension of particles in the air, it collects the scattered light at a certain angle and obtains a light dispersion curve that varies over time. The wavelength of the laser is 1100 μm. The particles of equivalent diameter and the number of particles with a different diameter per unit of volume can be calculated in real time by a microprocessor based on the MIE theory. In this case the sensor provides for PM10 and PM2.5 concentrations.

Since the ARPA data were compared with MVOBSV data that are based on the technology of laser scattering, a comparison with gravimetry was mandatory. The comparison between the two measurement techniques allowed for data validation. Figure III.2.3 reports the daily average PM10 and PM2.5 concentrations measured by the VC station with the two measurement techniques. From inspection of Figure III.2.3, it is clear that the values measured with gravimetry and laser scattering are almost the same, with an average difference of 0.23% for PM10 and 0.5% for PM2.5. The good agreement of the data measured means that the laser scattering measurements are reliable for further analysis. The MV station is placed on the roof 18 m from the ground, above the urban canyon, which acts as a mixing cell. This height should ensure that Saharan dust is not directly measured since it is mainly made up of PM10 and naturally tends to stagnate on the ground. The proposed approach consisted of the following fundamental steps:

- identification of the ARPA monitoring device with the same technological characteristics of the monitoring station to be compared, located at a distance, not exceeding 1000 m, and placed in the same urban context,
- validation of experimental data of basic pollution conditions through the analysis of comparison graphs,
- identification of Saharan events,
- comparison of the experimental daily values obtained by the measuring device with the ARPA values, and subsequently investigating in more depth the days with greater relevance using the hourly averages.

In this case study, the monitoring station analyzed was represented by MV and the ARPAC monitoring station used for the comparison was VC. The two stations considered (MV and VC) are both based on laser scattering technology and are located at a distance of 600 m in a context of medium urbanization.

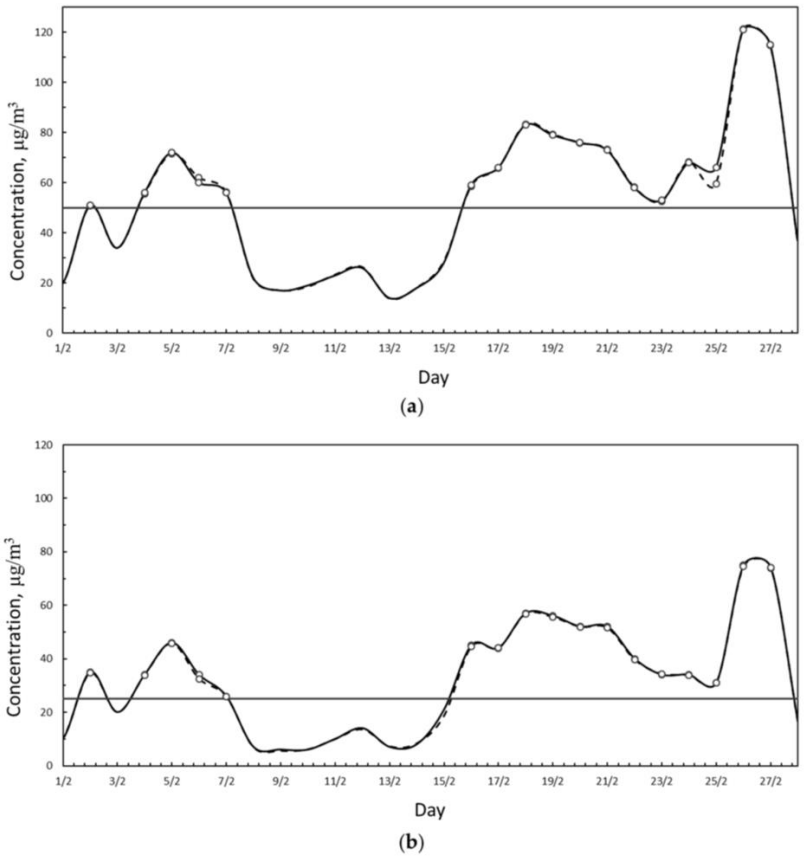


Figure III.2.3. *a) PM10 daily average concentration measured at the VC station with gravimetry (solid black line) and with laser scattering (dashed black line). b) PM2.5 daily average concentration measured at the VC station with gravimetry (solid black line) and with laser scattering (dashed black line). Constant grey line represents the legal limit concentration according to D.Lgs. 155/2010. Dots represent exceedance of the law limit.*

Meteorological data:

Meteorological data are fundamental for pollution dispersion analysis. The data used in this study were provided by an automatic weather station (Davis Vantage Pro 2) co-located with an air quality sensor. This station is part of the meteorological network managed by the non-profit organization MVOBSV and includes sensors for the measurement of essential atmospheric parameters, i.e., air temperature, relative humidity, air pressure, rainfall, and wind speed and direction. The data are stored with a temporal resolution of 10 min and are fully available for the analyzed period.

Results

In the first analysis, the daily data provided by the MV station was compared with those of the station VC which is closest (Figure III.2.4). Comparing the PM10 and PM2.5 trends of the stations showed a good agreement between the measurements. The measured data aligned around the central trend line, with small deviations, especially for PM2.5.

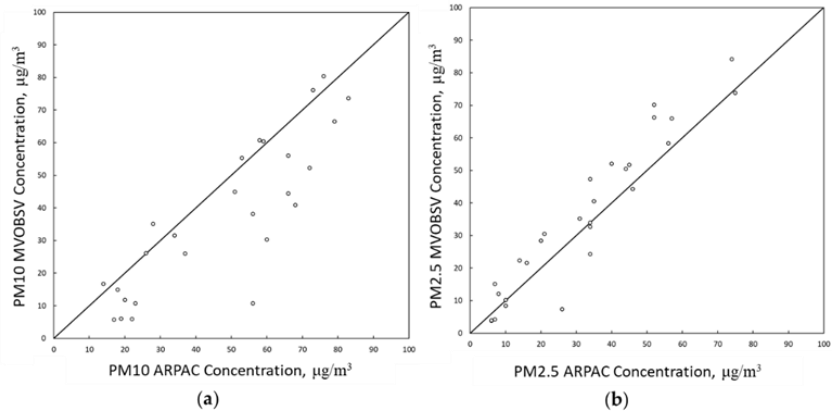


Figure III.2.4. *a) Comparison of PM10 and b) PM2.5 daily average concentrations measured at the MV and VC stations.*

From the analysis of the PM10 and PM2.5 trends (Figure III.2.5) it can be seen that for many days in February, both the recorded PM10 and PM2.5 exceeded the legal limit. In particular, the days exceeding the limit threshold of 50 µg/m^3 for PM10 were 16 at the SA station and 17 at the VC station. For PM2.5, the daily limit of 25 µg/m^3, similarly to PM10, was exceeded 16 times at the SA station and 17 times at the VC station. In February, the days when the pollution levels were low coincided with favorable weather conditions for dispersion. The three data trends have the same behavior, and differences can be attributed to the different positions in the urban context and, therefore, to the local conditions.

Subsequently, the trends of PM10, PM2.5, and their ratio were analyzed for each measuring station (Figure III.2.6). An inspection of Figure III.2.6 reveals that in the periods 5–7 and 23–28 of February, the difference between PM10 and PM2.5 concentrations was high for the SA and VC stations (Figure III.2.6 a,b); moreover, in such periods, the ratio between the particulate matter was low. This means that the contribution of PM10 to pollution was higher than PM2.5, and in particular, an external contribution represented by Saharan dust was added to the PM10 produced in the urban context. This external contribution was represented by the Saharan dust that

in these days was dispersed over Italian territory. The transported particulate is of desert and sandy origin, mainly composed of siliceous materials; unlike PM10 of urban origin which is characterized by the presence of carbon and metal-based particles. The analysis confirmed that the contribution of the mineral aerosol to the PM10 values can be very important, favoring the breaking of the legislative limit. Looking in depth at the two periods selected, it is possible to analyze the hourly trends of pollutants (Figures III.2.7 and III.2.8). In the first Saharan event, PM10 concentration reaches a peak of 180 g/m^3 measured at the VC station. As highlighted in Figure III.2.7, the PM10/PM2.5 ratio was extremely variable at the VC station (Figure III.2.7a), while it was slightly variable at the MV station (Figure III.2.7b). Moreover, at the MV station there was no evidence of Saharan dust contributing to PM10 concentrations, which remained proportional to PM2.5.

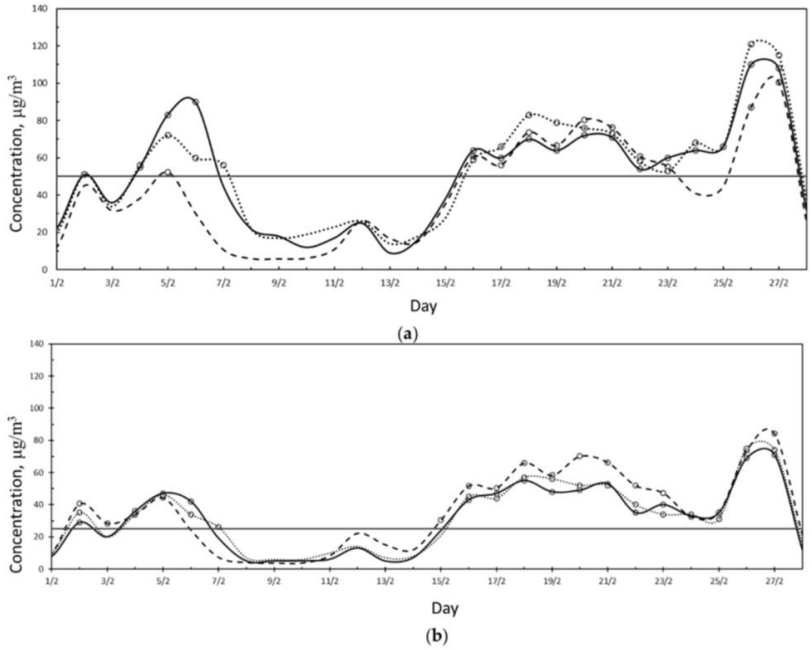


Figure III.2.5. *a) PM10 and b) PM2.5 daily average concentration measured at the SA station (solid black line), VC station (dotted black line), and MV station (dashed black line). Constant grey line represents the legal daily average limit concentration according to D.Lgs. 155/2010. Dots represent exceedance with respect to the legal daily average limit.*

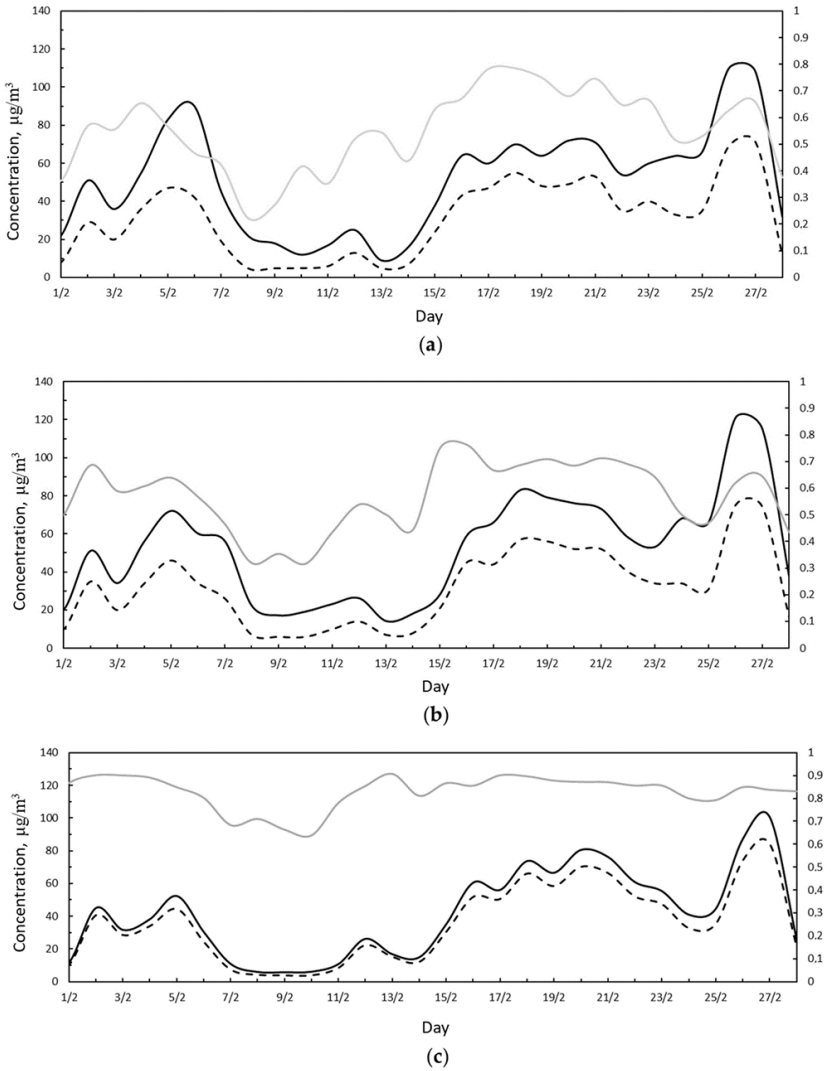


Figure III.2.6. *Daily PM10 (black solid line), PM2.5 (black dashed line), and PM2.5/PM10 concentrations (grey solid line) measured during February 2021 at the a) SA station, b) VC station, c) MV station.*

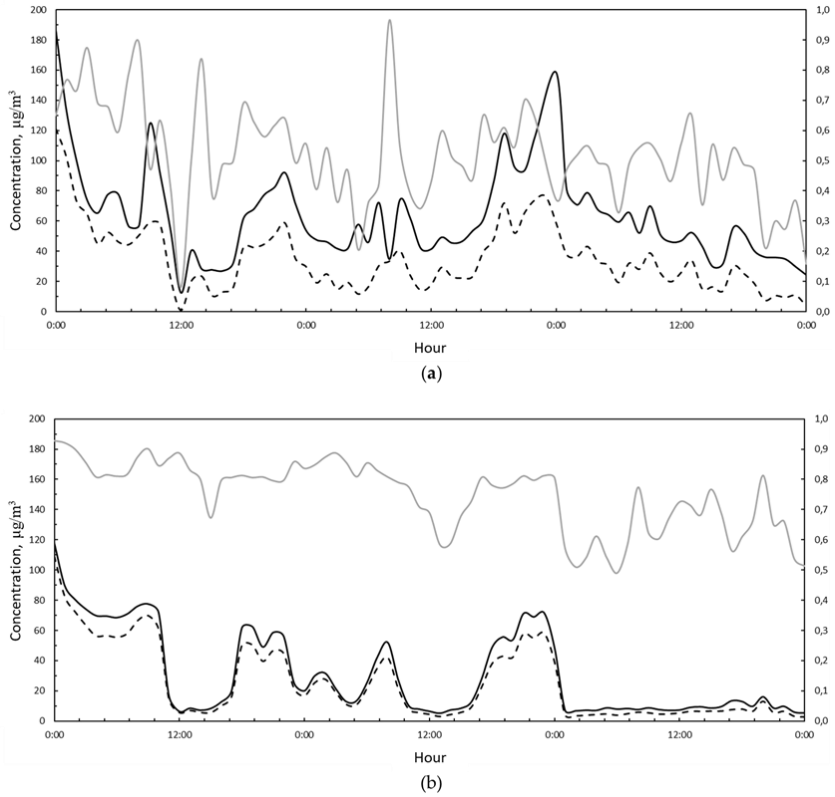


Figure III.2.7. *PM10 (black solid line), PM2.5 (black dashed line), and PM2.5/PM10 hourly concentrations (grey solid line) measured during the period 5–7 February 2021 at the a) VC station, b) MV station.*

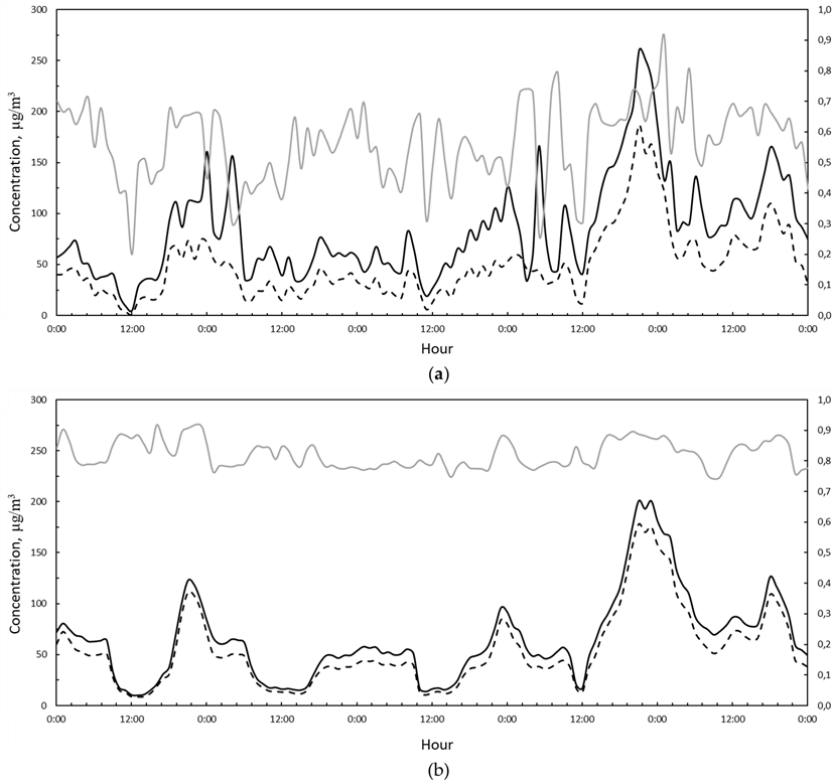


Figure III.2.8. *PM10 (black solid line), PM2.5 (black dashed line), and PM2.5/PM10 hourly concentrations (grey solid line) measured during the period 23–28 February 2021 at the a) VC station, b) MV station.*

Considering the second Saharan event, during 23–28 February and reported in Figure III.2.9, the PM10 concentration reached an hourly peak of 260 µg/m^3 measured by the VC station (Figure III.2.9a). As in the previous case, the MV measuring station did not show any evidence of an external PM10 concentration contribution, in fact, the PM10/PM2.5 ratio was slightly variable and the pollutant trends were similar (Figure III.2.9b). There was only a slight difference between PM10 and PM2.5 on days 26–27. The absence of evidence of a Saharan influence phenomenon at the MV station is probably due to the different height at which it is positioned compared to the ARPAC monitoring stations. The MV station is positioned on the roof of a building, while the two ARPAC stations are located near the street. The open-field exposure of the MV station does not allow the detection of the Saharan phenomenon, since the dust is more dispersed and diluted in the atmosphere at that height. Further analysis included the daily pollution inspection, to compare the trends measured by the VC and MV stations. In particular, some of most interesting days were more deeply analyzed, and

also including the meteorological parameters. In Figure III.2.10, the hourly measured concentration trends at the VC and MV stations appear to be in good agreement on most of the days analyzed. In detail, the measurements were also in agreement in their values and not only in their trend in 25% of cases.

On both days, at around 7 and 9 a.m. the pollutant levels increased. Particulate matter tends to settle on the ground due to gravity, both when airborne and as a result of rain. It remains on the asphalt if there is no movement of vehicles, such as during the night. When city actIIIities start early in the morning and vehicles start moving on the roads, the particulate matter is dislodged from the road surface and released back into the atmosphere, increasing particulate concentrations. This phenomenon was more evident at the MV station than the VC station because it is located in the more congested city center. Figure III.2.11 shows the 500-hPa and sea level pressure fields for a typical scenario in which low pollutants concentrations were detected in the area of Avellino due to the simultaneous action of rain and winds. More specifically, Figure 10a shows the synoptic scenario at 00:00 UTC on 11 February 2021, when the Italian peninsula was affected by a low-pressure area, well-structured at both mid and low tropospheric levels. This cyclonic system caused light rainfall in the study area (the AWS involved in this study registered a 24-h accumulated precipitation value of 1.5 mm), associated with moderate south-western winds. In the subsequent 12 h, the low-pressure area moved towards the Balkan peninsula (Figure III.2.11b); this caused an improvement in weather conditions in Avellino. However, a moderate breeze from the west was still observed and, therefore, favorable pollutant dispersion renewed.

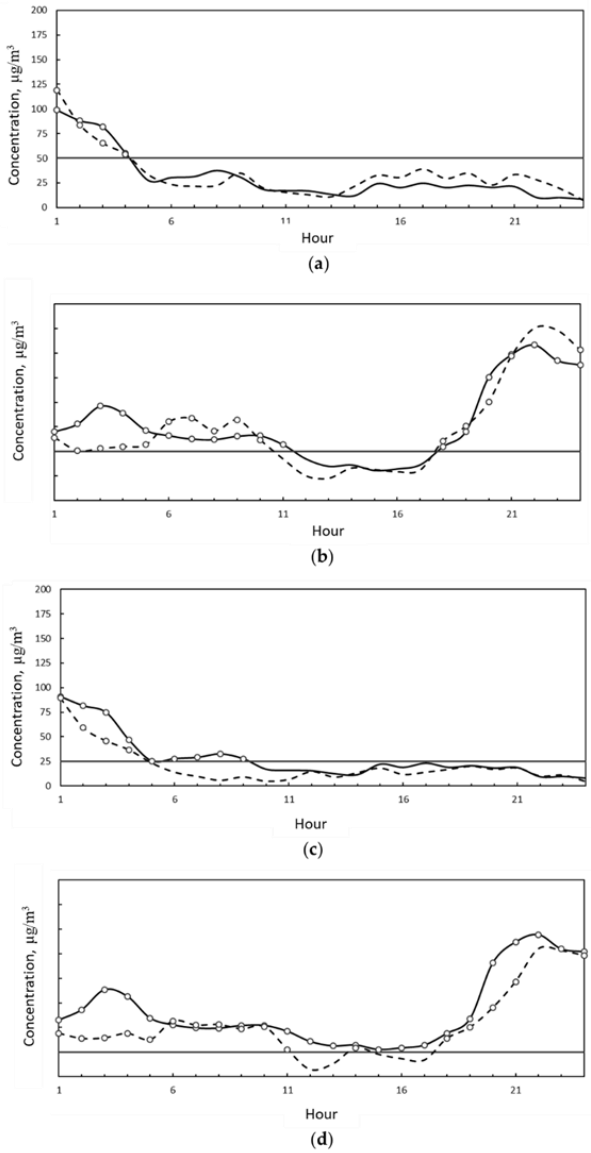
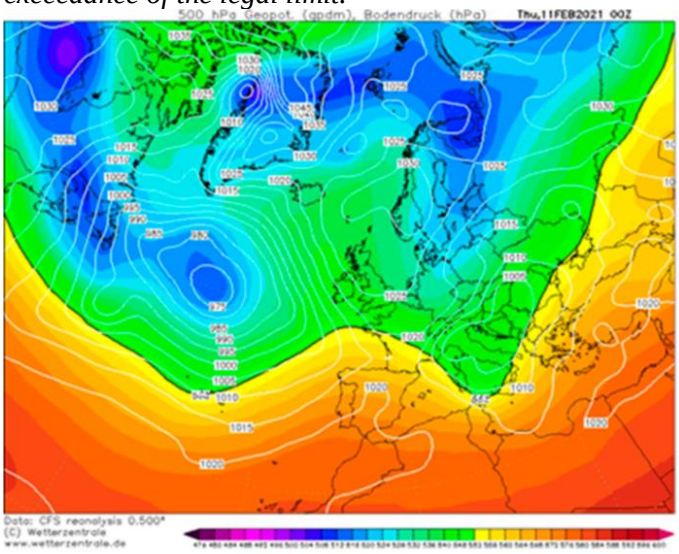
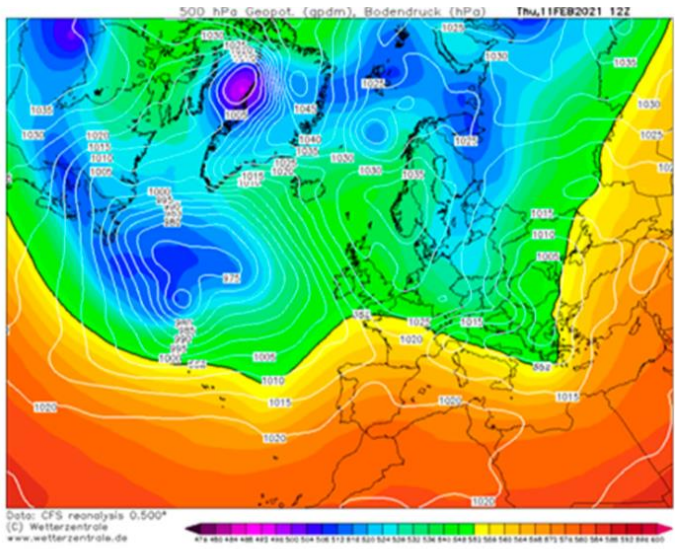


Figure III.2.9. *a) PM10 hourly average concentration measured at the VC station (solid black line) and MV station (dashed black line) 3 February 2021. (b) PM10 hourly average concentration measured at the VC station (solid black line) and MV station (dashed black line) 21 February 2021. (c) PM10 hourly average concentration measured at the VC station (solid black line) and MV station (dashed black line) 3 February 2021. (d) PM2.5 hourly average concentration measured at the VC station (solid black line) and MV station (dashed black line) 21 February 2021. Constant grey line represents*

the legal limit concentration according to D.Lgs. 155/2010. Dots represent exceedance of the legal limit.



(a)



(b)

Figure III.2.10. *a) Geopotential height at 500 hPa (shaded colors) and sea level pressure isobars (white solid lines) fields at 00:00 UTC on 11 February 2021. b) Geopotential height at 500 hPa (shaded colors) and sea*

level pressure (white solid lines) fields at 12:00 UTC on 11 February 2021 and 12:00 UTC. Both fields were retrieved from Climate Forecast System Renalysis of the National Centers for Environmental Prediction.

The Campania region was affected by two Saharan events in the periods 5–7 and 23–28 of February. Here, we provide a brief meteorological analysis of such events, and to better visualize the influence of Saharan dust on the Italian peninsula, synoptic maps were created.

On 5–7 February 2021, the European synoptic scenario resembled one of the most favorite patterns for the incoming Saharan air masses in the central Mediterranean area. The analysis of the 500-hPa geopotential height field revealed that, on 5 February (Figure III.2.12a), the presence of a trough over Western Europe extended from the British Isles to the western sector of Morocco. This trough was the result of a relevant oscillation of the polar front that occurred in the East Atlantic between 2 and 4 February. The sea-level pressure distribution (also depicted in Figure III.2.12a) showed a low-pressure area downstream of the trough axis, near the Gibraltar strait. A strong ridge, synonymous with stable weather conditions, instead modulated the atmospheric scenario in the Central and Eastern Mediterranean basins. In the subsequent 24 h, the trough moved eastwards (Figure III.2.12b) and caused a drop of atmospheric pressure on the western side of the Italian peninsula. The consequence of this pressure pattern was a dry and warm meridional flow, which advected on the central Mediterranean area subtropical continental air masses coming from the Sahara region. In this respect, the 850-hPa equivalent potential temperature field presented in Figure III.2.12c gives clear evidence of the warm tongue extending from the inland sectors of North Africa up to the Italian peninsula. On 7 February, the cyclonic system moved further to the east, causing a strengthening of warm air advection on southern Italy in the morning hours, and, consequently, a massive transport of Saharan dust. The meteorological conditions abruptly changed in the evening hours of 7 February, when the subtropical continental air was replaced by colder and moist air masses coming from the west.

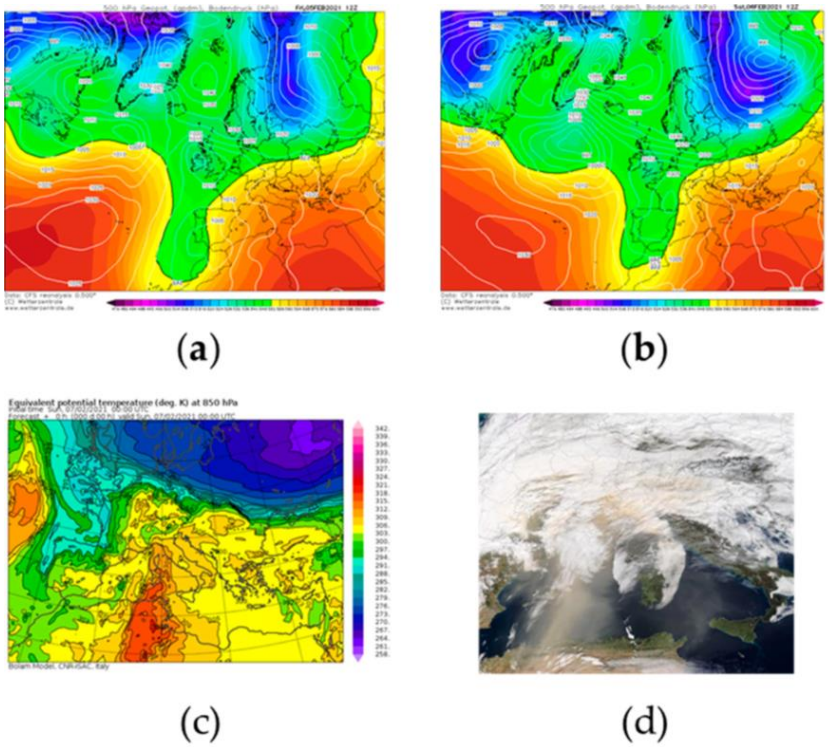


Figure III.2.11. *a) Geopotential height at 500 hPa (shaded colors) and sea level pressure isobars (white solid lines) fields at 12:00 UTC of 5 February 2021. (b) Geopotential height at 500 hPa (shaded colors) and sea level pressure isobars (white solid lines) fields at 12:00 UTC 6 February 2021. Both fields were retrieved from Climate Forecast System Renalysis of National Centers for Environmental Prediction. (c) 850 mb equivalent potential temperature (shaded colors) at 00:00 UTC 7 February 2021 from the BOLAM model of the Institute of Atmospheric Science and Climate of the Italian National Research Council. (d) Image of the Saharan dust event acquired on 6 February 2021 by the Moderate Resolution Imaging Spectroradiometer (MODIS) on board NASA's Terra satellite. Images in panels (a) and (b) are courtesy of http://www.wetterzentrale.de, (accessed on 23 February 2021) whereas the image in (c) was retrieved from https://www.isac.cnr.it/dinamica/projects/forecasts/index.html. (accessed on 23 February 2021) The picture in panel (d) is courtesy of the NASA Worldview Snapshots application (https://wvs.earthdata.nasa.gov) (accessed on 23 February 2021.*

Figure III.2.12d offers a true-color image of the central and western Mediterranean basins, collected by The Moderate Resolution Imaging Spectroradiometer (MODIS) on board NASA's Terra satellite on 6 February

2021. This picture highlights the Saharan dust cloud rising from North Africa and crossing the Mediterranean Sea. Saharan event occurring on 23–26 February 2021. A second relevant Saharan event took place in the study area at the end of February 2021 (days 23rd–26th). The meteorological dynamics that forced the incoming of desert dust in the central Mediterranean area had some similarities with those involved in the previous event. To shed light on these dynamics, it is necessary to analyze the 500-hPa geopotential height field on 22 February at 12 UTC (Figure III.2.13a), which reveals the presence of a cut-off low over the Alboran Sea, generated by a Rossby wave that affected the Iberian Peninsula and Morocco during 20 and 21 February. This upper-level circulation was associated with a low-level cyclonic area, located inland of Algeria, as clearly highlighted by the sea-level pressure field depicted in Figure III.2.13a. On the Italian peninsula, stable weather conditions prevailed, due to the presence of a ridge. This synoptic scenario promoted an advection of warm sub-tropical continental air, which can be easily identified by means of the 850-hPa equIIIalent potential temperature field (see Figure III.2.13b). The latter clearly shows the pattern followed by warm air, which reached central Europe after crossing the Sardinian Sea. On 23 February 2021, the ridge located over Italy strengthened (see Figure III.2.13c), obstructing the natural eastward movement of the low-pressure area located over the western Mediterranean basins. This evolution led to the translation of the cyclonic system inland of Algeria and, therefore, to the end of the warm advection in the central Mediterranean area. The previously advected sub-tropical air (rich in Saharan dust), following a clockwise motion imposed by the anticyclone, spread out over the Italian peninsula and Balkan regions. Owing to the persistence of the ridge in the subsequent days, the Saharan dust cloud continued to flow around Italy although it had been inevitably subjected to a gradual dilution process. Satellite evidence of Saharan dust affecting central and northern Italy is provided by the image in Figure III.2.12d, acquired by MODIS on 23 February 2021.

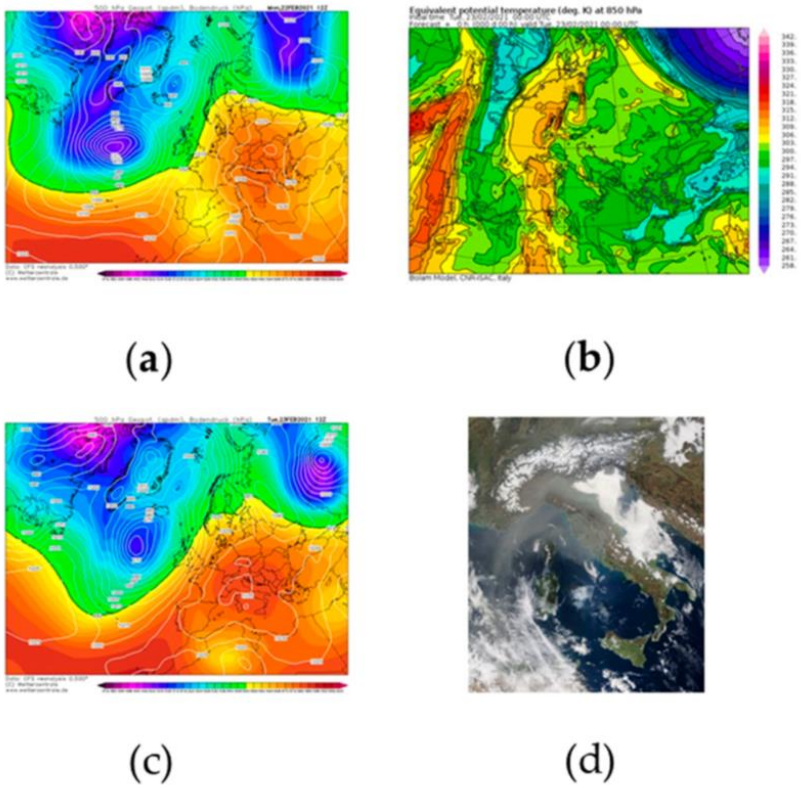


Figure III.2.12. *a) Geopotential height at 500 hPa (shaded color) and sea level pressure (white solid line) fields at 12:00 UTC of 22 February 2021. Fields were retrieved from Climate Forecast System Renalysis of National Centers for Environmental Prediction. (b) 850 mb equivalent potential temperature (shaded color) at 00:00 UTC of 23 February 2021 from the BOLAM model of the Institute of Atmospheric Science and Climate of the Italian National Research Council. (c) Geopotential height at 500 hPa (shaded color) and sea level pressure (white solid line) fields at 12:00 UTC of 23 February 2021. Fields were retrieved from Climate Forecast System Renalysis of National Centers for Environmental Prediction. (d) Image of the Saharan dust event acquired on 23 February 2021 by the Moderate Resolution Imaging Spectroradiometer (MODIS) on board NASA's Terra satellite. Images in panels (a) and (b) are courtesy of http://www.wetterzentrale.de (accessed on 23/02/2021, whereas the image in (c) was retrieved from https://www.isac.cnr.it/dinamica/projects/forecasts/index.html (accessed on 23/02/2021). The picture in panel (d) is courtesy of the NASA Worldview Snapshots application (https://wvs.earthdata.nasa.gov) (accessed on 23/02/2021).*

The analysis of Saharan dust contribution to PM10 in urban areas is defined by the high levels of PM10 measured not followed by a simultaneous increase in PM2.5 concentrations. By applying to this study case the method proposed by Escudero et al. it was possible to define the contribution of PM10 measured on the days in which the Saharan events occurred, using the 30th percentile of the monthly concentrations. By applying this procedure to the background VC station, we obtained the PM10 value in the absence of mineral aerosol. Comparing the values obtained after the application of the procedure suggested by Escudero et al., PM10 values similar to those measured at the MV station were obtained. As an example, for the VC station, a daily value of 60 $\mu g/m^3$ was measured, and the monthly 30th percentile value was 27 for this month. Therefore, the net dust contribution at this point, for this day, was 33 $\mu g/m^3$. For the same day considered, the MV station recorded a PM10 daily value of 31 $\mu g/m^3$. Comparing this last value with the previous one obtained without considering the mineral dust contribution to the PM10, it is clear that they are similar. This confirms that the Saharan dust phenomenon was not measured from the station located at a height of 15/20 m, as suggested by the method proposed in this study. With this approach it is possible to also evaluate the exceedance in PM10 concentrations due to the mineral dust to the urban background. This type of approach can be used in any city with orographic characteristics, urbanization, industrial supplies, population density, type of heating, and degree of traffic similar to the city in the case study. The peculiarity of being in a basin, surrounded by mountains is a frequent feature of Italian and European cities, which makes the case study adaptable to many real situations, such as the Po valley in Lombardy (Italy).

Conclusions

The occurrence of natural events such as the transport of Saharan dust can lead to a significant increase in the PM10 concentrations measured in urban areas. The correct analysis of the air quality data, including the evaluation of the influence of these natural phenomena, is necessary to define the real levels of urban pollution. The Saharan phenomena bring the levels of PM10 to very high values, as in this case study to 260 $\mu g/m^3$, which if interpreted incorrectly can generate unnecessary alarmism in the population and institutions. In fact, due to their nature, the events that transport Saharan dust are completely unpredictable and therefore unmanageable by man.

III.3 One-day forecast model for pollution intensity in a single position

The relationships between meteorological factors, traffic flow, topographical factors and particulate concentration have been widely analyzed in literature elsewhere using models with inferential statistics, such as linear regression or correlation analysis (Sahanavin et al., 2018). In this case, a new methodology is proposed able to forecast the urban quality air using the concentration levels of particulate matter and meteorological conditions. The air quality data recorded by monitoring stations installed in a southern city of Italy during the whole month of August 2018 were considered. From 1st September to 10th November, a comparison was made between the predicted concentrations of PM2.5 and PM10 using the OSPM model with the measured concentrations by monitoring stations to fine-tune the model. Finally, a correlation between air pollution and some meteorological factors (wind speed and direction, humidity) was investigated. The idea is to use the first 31 days to train the model and predict the 32nd and compare it with the one measured. Subsequently, the training part of the dataset will increase of one day, arriving to the last day.

Data collection

The concentration of small particulates (PM10 and PM2.5) and meteorological parameters (wind direction and velocity, temperature, humidity) have been collected for all days of August 2018 by three monitoring stations installed in Battipaglia city (Italy) in the position located between the industrial area and the urban center.

August-day data analysis

Provided the data at every hour, the data processing returns average values, $\bar{C}_{i,j,k}$ in Eq. III.3.1. Data were collected by value intervals for each different parameter (relative humidity, wind intensity and intensity, temperature, daily hour). The raw data were examined and some adjustments were made to purify data by outlier values. The time variability of data was considered studying its behavior during 24 hours of a day.

PM concentration values by the forecasting model

These data have been used to find the best fitting parameters of the model. Initial weight was considered based on the magnitude of the

difference between 75° and 25° percentiles (box in Figure III.3.2). So, for each parameter, the variability of data was evaluated considering the PM2.5/PM10 mean concentration in the parameter interval. A deterministic approach was used. In Eq. III.3.1, $c_{j,h}$ is the forecasted concentration for the day j and hour h, $p_{i,j-1}$ is the model coefficient for the parameter i optimized by day j-1, $\bar{c}_{i,j-1,k_{i,j,h}}$ is the average concentration (diamond in Figure III.3.2) considering only parameter *i*, until day *j-1*, in the parameter interval (by Figure III.3.2) $k_{i,j,h}$, $k_{i,j,h}$ is the interval related to parameter *i* from forecasting database for the day *j* for the daily hour *h*.

$$c_{i,h} = \sum_{i=1}^{n_p} p_{i,j-1} c_{i,j-1,k_{i,j,h}} \quad \text{(III.3.1)}$$

It has been imposed that the parameters coefficients to be estimated must have values between 0 and 1. In fact, it is assumed that the concentration of PM10 and PM2.5 is composed of all positive contributions and that they can have as a maximum value the sum of the contributions of the individual components. A further plausible constraint could be that the sum of the coefficients is equal to 1 but this hypothesis has not yet been taken into consideration. In previous works the coefficients have been evaluated as a proportion of the amplitude of the interval P0.75 and P0.25 so that the sum of these coefficients equal to 1. In this work these values have been used as initial values of the objective function. The objective function is robust with respect to the choice of the initial values, in fact the same solutions are reached even with different percentiles or random values between 0 and 1. It is evident that the function in the identified domain shows a single minimum.

Objective function description

To find the best value of the coefficient associated with each parameter, an objective function was described as the sum of the quadratic deviations of the residuals between the estimated and the measured value. In Eq. III.3.5., *n* is the day considered (from 1st to 71st), *h* is the daily hour, $c'_{j,h}$ is measured PM concentration by the monitoring stations. The coefficient estimation starts the first day using all the background of the previous 31 days of August. Going forward in the estimation of the coefficients, the number of days used as background for the estimation will progressively increase. In this way, the minimization takes into account not only what happened in the last 24 hours, but also in the previous n days considered.

$$f_{OB,n} = \sum_{j=1}^{n} \sum_{h=1}^{24} \left(c_{j,h}^{\cdot} - c_{j,h}\right)^2 \quad \text{(III.3.2)}$$

Model tuning

From the 1st September to 10th November 2018 (71 days) a comparison was made between the predicted concentrations of PM2.5 and PM10 using the dispersion model with the measured concentrations by monitoring stations in order to fine-tune the model. It was found that after n_{TOT} (71) days, a reliable model was obtained and no further adjustment was required to adjust prediction with experiments (Fig. III.3.1).

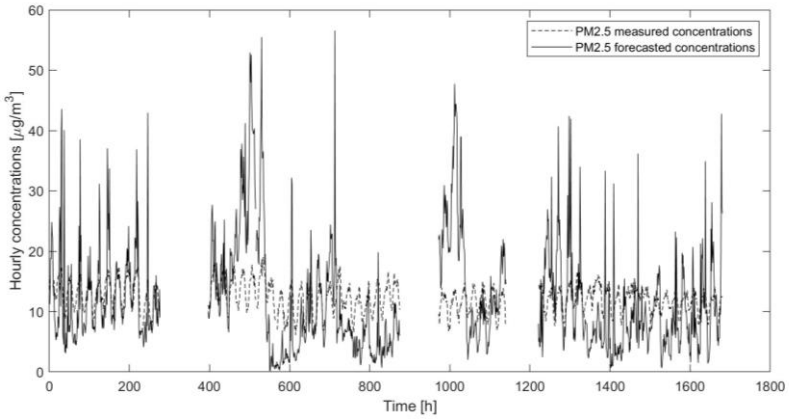


Figure III.3.1. *PM2.5 concentrations forecasted and measured during the time interval of 71 days.*

Equations III.3.1 and III.3.2 allow to estimate the parameters coefficients in an iterative way through a numerical optimization (minimization) algorithm. Equation III.3.1 can be related to a polynomial regression as described by Equation III.3.3:

$$y = X\beta + \varepsilon \quad \text{(III.3.3)}$$

where

y is the vector that contains $c_{j,h}$;
X is the PM concentration matrix for each parameter
β is the coefficient to be estimated

The polynomial regression implemented leads to an analytical estimate of the coefficients through OLS (ordinary least squares estimation) as Equation III.3.4.

$$\beta = (X^T X)^{-1} X^T Y \quad \text{(III.3.4)}$$

The coefficients thus estimated can assume all possible values in the domain -∞, +∞. Given the previously clarification, it is a constrained estimation with coefficients that can only assume values in [0, 1]. For this reason, an estimate through a numerical optimization algorithm was preferred instead of the analytical evaluation.

Results

On each box in Figure III.3.2, the central mark indicates the median, the diamond indicates the mean value and the bottom and top edges of the box indicate the 25th and 75th percentiles, respectively. Points are considered as outliers, indicated by a cross. Figure III.3.2 shows clearly that the PM concentration varies in different ways on each parameter. The most significant variation of concentration depends on relative humidity (*(a)* in Figure III.3.2), wind intensity (*(b)* in Fig.III.3.2) and hour (*(e)* in Fig.III.3.2). The data used for the boxplot in Figure III.3.2 represents the S1 station. The trends inside the boxplot for S1 are similar for the ones obtained for the stations S2 and S3. For this reason, only the boxplot for S1 is reported.

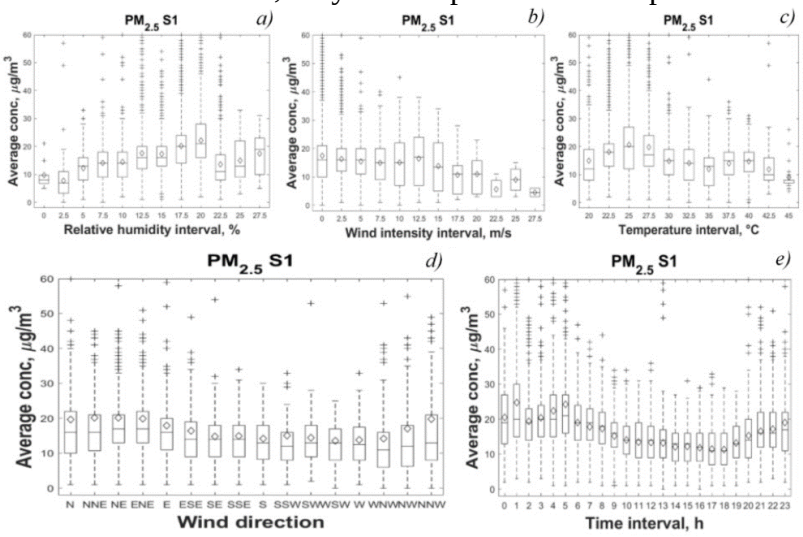


Figure III.3.2. *Variation of the average concentration of PM2.5 for S1 respect to intervals of a) relative humidity; b) wind intensity; c) temperature; d) wind direction; e) daily time. Data processing was carried out between*

1st-31st August. Central mark indicates the median. The bottom and top edges of the box indicate the 25th and 75th percentiles, respectively. Diamonds indicate the mean value. Outliers are plotted individually using the '+' marker symbol.

In order to estimate the robustness of the model, the RMSE (Root-Mean-Square Error) values were evaluated every day as:

$$RMSE_j = \sqrt{\frac{1}{24} * \sum_{h=1}^{24} \left(c_{j,h}^{\cdot} - c_{j,h}\right)^2} \quad \text{(III.3.5)}$$

Finally, to summarize the performance of the model for each station, the average RMSE was calculated by:

$$\overline{RMSE} = \sum_{j=1}^{n_{TOT}} RMSE_j \quad \text{(III.3.6)}$$

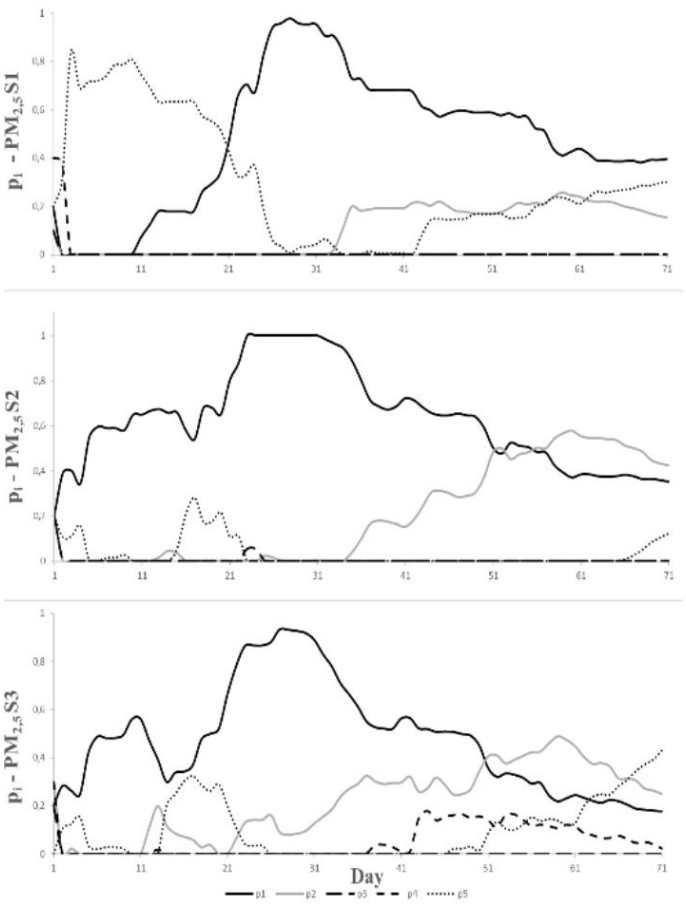


Figure III.3.3. *Model coefficients pi from 1st to 71st day for PM2.5 concentrations for stations S1, S2, S3.*

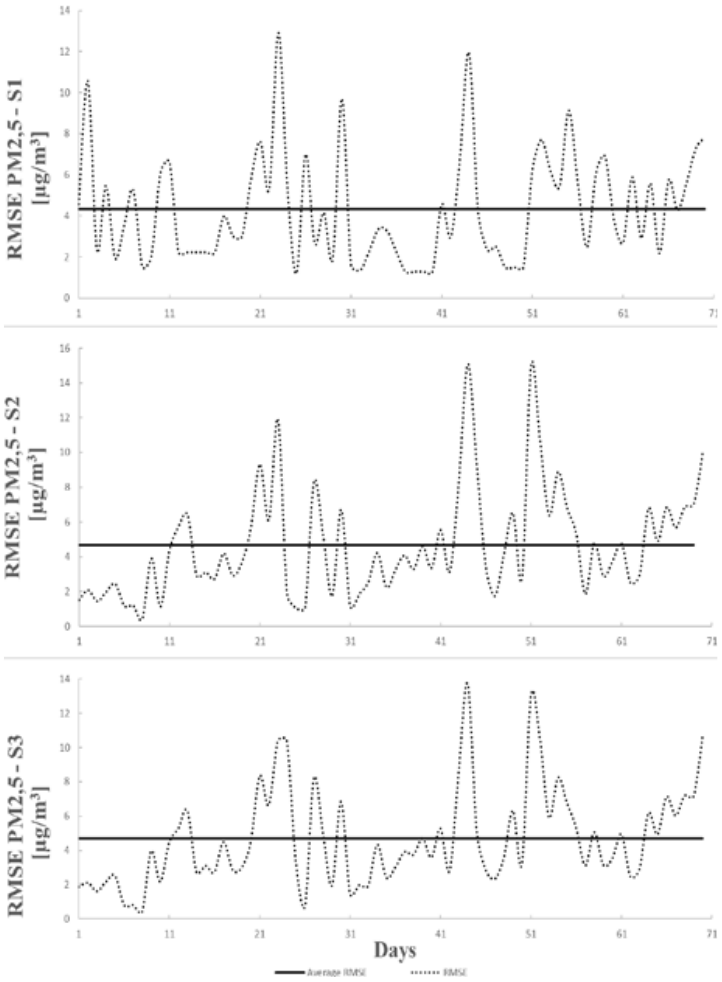


Figure III.3.4. *RMSE values from 1st to 71st day for PM2.5 concentrations for S1, S2, S3. Black solid line for Average RMSE and black dotted line for RMSE.*

Figure III.3.3 shows the behaviour of the model coefficients for each day used for the minimization. The coefficients are more varying in the first days, while they are taking around a constant value after 40-50 days of observation (Table III.3.1). The coefficient estimation is also essential to understand what parameter influences the phenomenon studied. As a result, relative humidity, wind intensity and time are the parameters that influence mainly the PM concentrations. This evidence is also underlined in Figure III.3.3, where PM concentrations change during the day above all, according to the anthropic activities. The PM concentrations decrease by increasing the

wind intensity because of air origin from the sea (SE direction) or mountains (other directions).

Table III.3.1. *Coefficients pi for the last day for PM2.5 and PM10 and stations S1, S2, S3.*

		PM2.5			PM10		
	Associated variable	**S1**	**S2**	**S3**	**S1**	**S2**	**S3**
p_1	Relative humidity	0.35	0.40	0.16	0.40	0.35	0.17
p_2	Wind speed	0.13	0.43	0.23	0.15	0.42	0.25
p_3	Temperature	0	0	0	0	0	0
p_4	Wind direction	0	0	0.01	0	0	0.02
p_5	Hour	0.27	0.1	0.41	0.3	0.12	0.43

RMSE values from 1st to 71st day are reported in Figure III.3.4. The dotted line represents the RMSE of Eq. III.3.3, the continuous black line is the average RMSE of Eq. III.3.4. In the early days, RMSE was low because the meteorological conditions were similar to the previous days on which the model parameters were based. Between the 23rd and the 31st day and between the 43rd and the 51st day the average error is very high (up to 15 $\mu g/m^3$ difference) due to sudden and robust weather variations. After the 51st day and until the end of the comparison the values appeared to be close to the average RMSE values of Table III.3.1. This demonstrates the robustness of the model. Table III.3.2 shows how the error values are lower (about 1 $\mu g/m^3$) for PM2.5 compared to PM10 due to lower concentration values for PM2.5.

Table III.3.2. *Average RMSE ($\mu g/m^3$) values obtained after 71st days for PM2.5 and PM10 concentrations for S1.*

PM2.5			PM10		
S1	**S2**	**S3**	**S1**	**S2**	**S3**
4.33	4.67	4.70	4.7	5.02	5.21

Conclusions

This part of the work was focused on the development of a forecast model that uses experimental data of concentration of particular matter PM2.5 and PM10. The data was obtained through 3 outdoor positioned monitoring stations in an area of about 10 km^2 between an industrial zone and a densely popular urban center. The model results show that the coefficients of the linear model are different after 71 days of optimization of the model. The coefficients are different from station by station. The more influence meteorological parameters are the relative humidity, wind speed

and hour. After 71 days, the model shows a sufficient agreement model-experimental data, equal to about 4 $\mu g/m^3$ is for PM2.5 and PM10. This value appears reasonable considering the considerable variability of concentrations of pollutants and that the limit value of the law is equal to 50 $\mu g/m^3$ for PM10 and 25 $\mu g/m^3$ for PM2.5. RMSE values don't improve with more days. This is due to the change of season influencing the data strongly due to the different correlations between weather and pollution. Further data (until 12 months) are necessary in order to make model able to fit very well pollution data for every season.

III.4 Pollution dispersion from a fire using a Gaussian Plume model

Gaussian Plume model

The Gaussian models have been the reference tools for the studies of the pollutant diffusion into the atmosphere and still continue to be valid supports for the assessment of emission scenarios. Their simplicity and immediacy of use and the reliability of the solutions adopted to model the effects of interaction of the plume emitted with the external environment, makes them the most used in the description of pollutant emissions. The emission description within the area studied takes place by associating to each meteorological condition inserted in the Gaussian plume input (values of wind speed and direction, temperature and atmospheric stability) whose shape is modulated by atmospheric conditions. The ease of use of the Gaussian models is linked to the fact that the model requires a type of standard data easily available in the literature both as regards the structural aspect of the emissive sources and the meteorological or geophysical one. The immediacy of use is linked to the rapid response of the model in returning the concentration results. The Gaussian model, implementing an already defined solution of the diffusion Equation, does not require its numerical resolution, thus resulting in extremely fast execution even of rather long time intervals.

The simplicity and immediacy of use in addition to the conservative parametric choices used in its calculation algorithms make it a very effective screening tool applicable to the analysis of the diffusion of non-reactive pollutants. Finally, since the quality of the results provided by the Gaussian models is closely linked to compliance with the conditions of applicability and the quality of the meteorological data that are used as input, compliance with these characteristics makes them very effective tools both in the quantitative control phase and as analysis and interpretation tool for the results obtained. The Gaussian plume dispersion is therefore, the one that better describes the case studied in this work. The fire can be similar to the pollutant emissions from a punctual source.

The application of the model includes some implementation steps that lead to the final results. The definition of the diffusion Equation and the boundary conditions are the first implementation step. Then, since the diffusion Equation includes dispersion parameters that depends on air stability conditions. So, it is necessary the definition of the wind field characteristics like wind intensity and wind stability class. Then, knowing the height of the stack and the position of the source, the concentration in the point measured, it is possible to obtain the fire mass flow rate.

The Gaussian dispersion model approach was adopted in this study for its simplicity and ease of application.

In order to describe the transport of particulates, which are scarcely affected by chemical transformations in the atmosphere, the advection-diffusion Equation describing the mass balance of pollutants in atmosphere does not consider any reaction term (Vesilind et al., 1994):

$$\frac{\partial C}{\partial t} + \frac{\partial uC}{\partial x} + \frac{\partial vC}{\partial y} + \frac{\partial wC}{\partial z} = K_x \frac{\partial^2 C}{\partial x^2} + K_y \frac{\partial^2 C}{\partial y^2} + K_z \frac{\partial^2 C}{\partial z^2} \quad \text{(III.4.1)}$$

where *C* is the pollutant concentration, *t* is time, *x* is the horizontal spatial coordinate in the wind direction, *y* is the horizontal spatial coordinate orthogonal to the wind direction, *z* is the vertical spatial coordinate, *u* and *w* are the horizontal and vertical components of the wind velocity, *Kx*, *Ky* and *Kz* are the dispersion coefficients in the three coordinate direction.

$$u \frac{\partial C}{\partial x} = K_y \frac{\partial^2 C}{\partial y^2} + K_z \frac{\partial^2 C}{\partial z^2} \quad \text{(III.4.2)}$$

Equation III.4.1 can be solved with the boundary conditions of a continuous point source and the solution is the following (Sutton et al., 1932):

$$C(x,y,z) = \frac{Q}{2\pi u \sigma_y \sigma_z} \exp\left(-\frac{y^2}{2\sigma_y^2}\right)\left[\exp\left(-\frac{(z-H)^2}{2\sigma_z^2}\right) + \exp\left(-\frac{(z+H)^2}{2\sigma_z^2}\right)\right] \quad \text{(III.4.3)}$$

Where *Q* is the pollutant mass flow rate coming out of the source, *H* is the source height above the ground and the mirror source term (a virtual source located at a height -H) is introduced to model the worst possible case of air pollution corresponding to a reflection boundary conditions on the ground. σy and σz are defined according the Pasquill-Gifford correlation (Gifford 1961):

$$\sigma_y^2 = \frac{2K_y x}{u} \qquad \sigma_z^2 = \frac{2K_z x}{u} \quad \text{(III.4.4)}$$

i.e. the width and the thickness of the plume described by σ_y and σ_z, both increase proportionally with the space time (x⁄u), that is the time taken by the pollutant to reach a certain position x since it was emitted from the source point. The values of the proportionality constants *Ky* and *Kz* between the space time and the standard deviation, depend on the tendency of the atmosphere to repress vertical convection, the so-called atmospheric stability. In fact, according to the model, the atmosphere stability affects the

parameters *Ky* and *Kz* so that unstable conditions have standard deviations that more rapidly increase in the wind direction while stable conditions have standard deviations that less rapidly in the wind direction. Hence, in stable conditions the pollutant may travel longer distances before dispersing. Figure III.4.1 illustrates the situation being modelled.

Figure III.4.1. *View of the area around Avellino (Italy) with the fire position (F) and the fixed monitoring stations (S1, A1, A2).*

At a certain downwind distance *x*, the maximum concentration lies on the plume axis of symmetry and is *Cmax*:

$$C_{max}(x) = \frac{Q}{2\pi\sigma_y\sigma_z u} \tag{III.4.5}$$

Therefore, the cloud smoke concentration downwind the stack and at the ground level is calculated as follows:

$$C(x,0,0) = \frac{Q}{\pi\sigma_y\sigma_z u} exp\left(-\frac{H^2}{2\sigma_z^2}\right) \tag{III.4.6}$$

The implementation of the Gaussian Plume model was obtained using a MATLAB code derived from the one freely available from the University of Manchester (Connolly P.). The code allows to customize several parameters to combine in the model more stack such the one reported above in Equation III.4.6 by defining the different stack positions on the ground plane, called stack *x*, stack *y*, the mass flow rate *Q* and the height of the stack *H*.

The values of the vertical stability parameter were set according to the Pasquill atmospheric stability classes (Ferrero. et al., 2019), which is the most commonly used method of classification of the turbulence in the atmosphere. According to this classification, there are six stability classes

named A, B, C, D, E and F with class A being the most unstable or most turbulent class, and class F the most stable or least turbulent class. Each class is defined by meteorological conditions that are the surface wind speed, daytime incoming solar radiation and, night-time cloud cover. Solar radiation increases atmospheric instability through warming of the Earth surface so that warm air finds itself below cooler (and therefore denser) air, a situation that promotes vertical mixing. Clear nights push conditions towards stable as the ground cools faster establishing more stable conditions and inversions. Wind increases vertical mixing, breaking down any type of stratification and pushing the stability class towards neutral (D) (Pasquill F., 1961). Incoming solar radiation is classified as follows: strong (> 700 W/m^2), moderate (350-700 W/m^2), slight (< 350 W/m^2) (Pasquill F., 1961).

To carry out a pollution dispersion analysis, firstly it is necessary to investigate the effects that the assumptions about wind direction have on dispersion of pollutants. Usually, the wind speed and the wind direction would be fed to the model after being determined by either observational data or from weather forecast. Then, it is possible to generate a synthetic dataset by either: (1) having the wind come from a constant direction; (2) having the wind come from a completely random direction and (3) having the wind come from a prevailing direction, with some variation either side.
In the model in this study, a different approach was followed to describe the change of the standard deviations describing the plume thickness and width.

$$\sigma_y = a * x^b \tag{III.4.7}$$

$$\sigma_z = c * x^d + f \tag{III.4.8}$$

where the *a*, *b*, *c*, *d*, *e* and *f* constants depend on the stability class and on the distance, i.e. they may be different for *x* values smaller or greater than 1 km (Vismara R., 1983).

The case study

In this work the migration of pollutants from a real accidental fire occurred in Avellino on 13/09/2019 was studied with the dispersion stack model. The fire broke out in a company producing plastic containers for automotive batteries. A small part of polypropylene plastic components and many wooden pallets were burned. The starting point of the fire is inside the Avellino industrial development area (F in Figure III.4.1).

Aerosol particles originating from wood burning are predominantly in fine particle mode (PM2.5) (Seinfeld et al., 2006) while the combustion of PP determines the presence of Pb, Cd and, Zn (Artaxo. et al., 2002) in the fire smoke. In the model, however, it is assumed that the fire smoke is made of CO_2 for 95% in mass and of PM2.5 for the remaining 5% and the corresponding value of the emission factor e is assumed equal to 0.05

(Meharg et al., 1995). In fact, PM2.5 generally includes all the metals delivered by smoke fires. Concentrations of PM2.5 was measured by a fixed monitoring station (S1 in Figure III.4.1) located at 4 km South-West from the fire. This concentration data was used to evaluate the mass flow of the smoke cloud at its origin in the fire. In Avellino there are also two fixed monitoring stations (A1 and A2 in Figure III.4.1) of the Regional Agency for the Environmental Protection (ARPAC) network providing data of the daily average concentrations. These data were used to verify the model predictions.

Results

Time-series of the PM2.5 concentrations measured by the S1 station are reported in Figure III.4.2. Inspection of the Figure reveals a neat increase of PM2.5 after six hours from the fire start (at 1.45 pm), showing a peak of 41 $\mu g/m^3$. During the night, from 18h to 6h of the next day the PM2.5 concentration fluctuates between 41-12 $\mu g/m^3$. The instantaneous PM2.5 values measured in the S1 station exceeded the threshold of 25 $\mu g/m^3$. However, according to Table III.4.1, the regulatory daily average limit, set to the same value by the Italian Law (D.Lgs. 155/2010), was never exceeded.

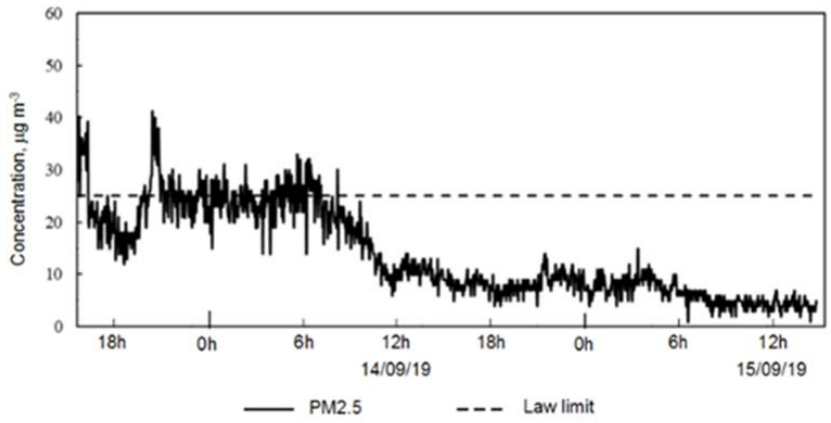


Figure III.4.2. *Instantaneous concentrations of PM2.5 measured by S1 and PM2.5. Law limit follows the Italian Law, DLgs. 155/2010.*

The PM2.5 concentration measured in real-time by the air quality monitoring station S1 after 12 hours was used as a reference value to infer the fire mass flow rate by applying the Gaussian Plume dispersion model. Subsequently, the best fitting value of the mass flow rate was used as an input to derive a more complete concentration map at the ground for the

Gaussian Plume dispersion model. The resulting model prediction values were compared with the values measured by the monitoring station A1 and A2. Then, to investigate the effect of the wind speed and the wind stability class on the pollutant dispersion, the Gaussian Plume model was applied for different conditions.

Table III.4.1. *Daily average concentrations measured from 13/09/2019 to 15/09/2019 in A1, A2, S1.*

	PM2.5 daily average concentrations μg / m³		
Station\date	**13/09/2019**	**14/09/2019**	**15/09/2019**
Arpac - Sc. V Circolo (A1)	19.0	13.5	11
Arpac - Scuola Alighieri (A2)	17.6	10	10.4
Sense Square (S1)	23.7	14.5	5.5

In particular, the data used to calculate the average value after 12 hours are provided by the monitoring station every 3 minutes. Data analysis of the distribution of measured values revealed that the interquartile range is 6 $\mu g/m^3$ (the first quartile is 21 $\mu g/m^3$, the median is 24 $\mu g/m^3$, the third quartile is 27 $\mu g/m^3$). As a result, due to the narrow range and the symmetric distribution during the 12 hours interval, the reference average value for the model was set equal to the median value of 24 $\mu g/m^3$.

To evaluate the mass of each of the two combustible materials burned, aerial pictures of the fire site were analysed. In fact, the fire occurred in the open-air deposit of the factory, where the materials were stocked. The number of wood pallets stocked in the area investigated was 1300. Considering that the weight of a single standard Epal pallet (UNI-EN 13698-1) is 25 kg, the total weight of wood pallet burned was $3.2 \cdot 10^4$ kg.

The amount of PP stocked was evaluated from the area occupied by the batteries containers that equals to about 2200 m^2. Considering a pile height of about 3 m, and the bulk density of the PP around 906 kg/m^3 and a void fraction of 0.9, the total mass of the PP stocked is of $5.0 \cdot 10^5$ kg. As a result, a total of $5.3 \cdot 10^5$ kg of fuel burned during the fire for a period of about 12 h, corresponding to an average rate of 12 kg/s. The equivalent stack height of the plume H was estimated equal to 75m, from visual inspection of pictures taken of the smoke column above the fire.

During the fire the wind was blowing from North East and, therefore, moving towards South-West (225°), at 7 m/s. Therefore, in the case study, the position of the air sampling station S1, placed at 4000 m from the fire was on the northern side of the developed smoke plume at about 500 m from the plume axis.

The experimentally measured mean concentration of 24 $\mu g/m^3$ was used in Equation III.4.6 to infer the mass flow rate of the pollutant to be used

within the model. Considering that the latter value of particulate concentration was mainly measured during the night and that the measured wind speed was about 7 m/s, the D stability class was assumed as a reasonable guess. A first order of magnitude estimate of the flow rate, to be used as a first guess for the subsequent steps was obtained by using Equation III.4.6. Afterwards, the best fitting mass flow rate value was obtained by minimizing the discrepancy between the measured value at the monitoring station S1. For this purpose, a numerical optimization algorithm was adopted. It was introduced a fixed orthogonal space reference system oriented with the ξ axis in the north direction and the η axis in the east direction and with the origin at the fire place. Following this reference system, the coordinates of the S1 monitoring station are (4000, 2200).

The mass flow rate producing the best fitting concentration at the monitoring station is 665 g/s. Comparing this value with the burning rate of 12 kg/s above reported, we estimate an emission factor to particulate matter e slightly larger than 5%. This value agrees with literature data (Meharg et al., 1995).

Figure III.4.3 reports the model results in terms of the distribution at the ground level of PM2.5 concentrations. At the monitoring station's position S1, the PM2.5 concentration evaluated by the model is 24.36 $\mu g/m^3$, with an error of 1.5% respect to the measured value of 24 $\mu g/m^3$. The maximum of the dust concentration is 950 $\mu g/m^3$, and it is found at a distance of about 1800 m downwind the fire. Considering the other experimental measurement points (A1 and A2 in Figure 5.4.1), it can be seen that the values estimated by the model are in agreement with those measured (average concentration of PM2.5 calculated as for station S1 and described above). In particular, at points A1 and A2, the estimated PM2.5 concentration value differs from the measured one (19.6 $\mu g/m^3$ for A1 and 14.1 $\mu g/m^3$ for A2) by 1.5% and 5.2%, respectively. The consistency of the experimental measurements with the values estimated by the model represents a further confirmation of the assumptions' validity.

Figure III.4.4 reports the ground level concentrations of PM2.5 estimated with the model in other stability classes. These provide a picture of what might have happened in different times of the day. It appears that, for increasing air stability, the dispersion along the distance decreases, the ground level values tend to increase and the position with the highest pollutant concentration moves away from the fire. The interdependent play between decreased vertical and horizontal dispersion indicated that the conditions corresponding to an intermediate stability class would have provided the highest pollution concentration at the sampling station position.
In order to assess the effect of the fire under different possible meteorological scenarios, by keeping constant the fire rate, the maximum PM2.5 concentrations and their corresponding positions were calculated for

all three air stability conditions (A, C, E) and with different wind speeds of 3 and 15 m/s.

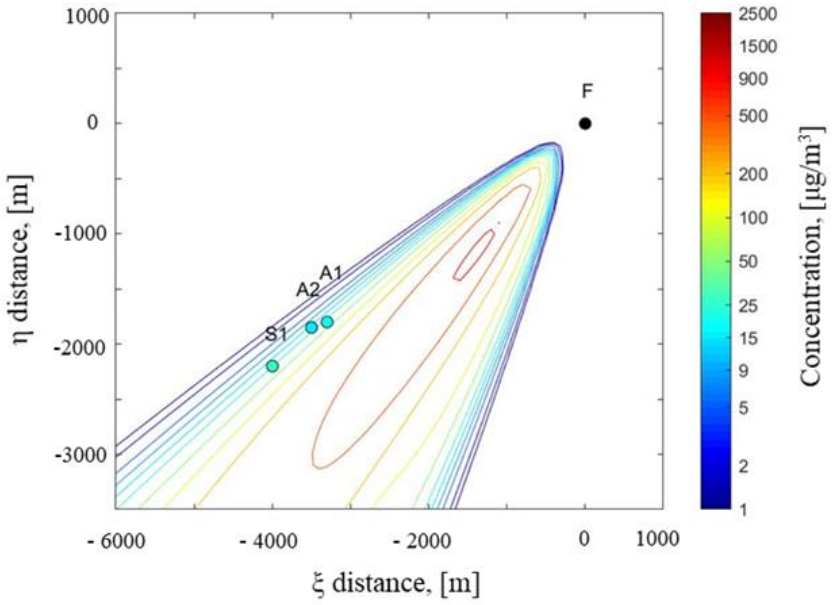


Figure III.4.3. *PM2.5 concentrations at the ground level according to the Gaussian model for the neutral stability class.*

Table III.4.2. *Maximum PM2.5 concentrations and corresponding position at variable wind speed and air stability.*

WIND SPEED m/s	**3**		**7**		**15**	
Stability class	max μg/m^3	x_{max} m	max μg/m^3	x_{max} m	max μg/m^3	x_{max} m
A	3.4e3	400	1.5e3	400	0.7e3	400
B	3.4e3	600	1.4e3	600	6.7e2	400
C	3.3e3	1000	1.4e3	1000	6.5e2	1000
D	2.2e3	1800	9.5e2	1800	4.4e2	1800
E	1.2e3	3600	0.5e2	3600	0.2e2	3600
F	0.5e2	6000	0.2e2	6000	0.1e2	6000

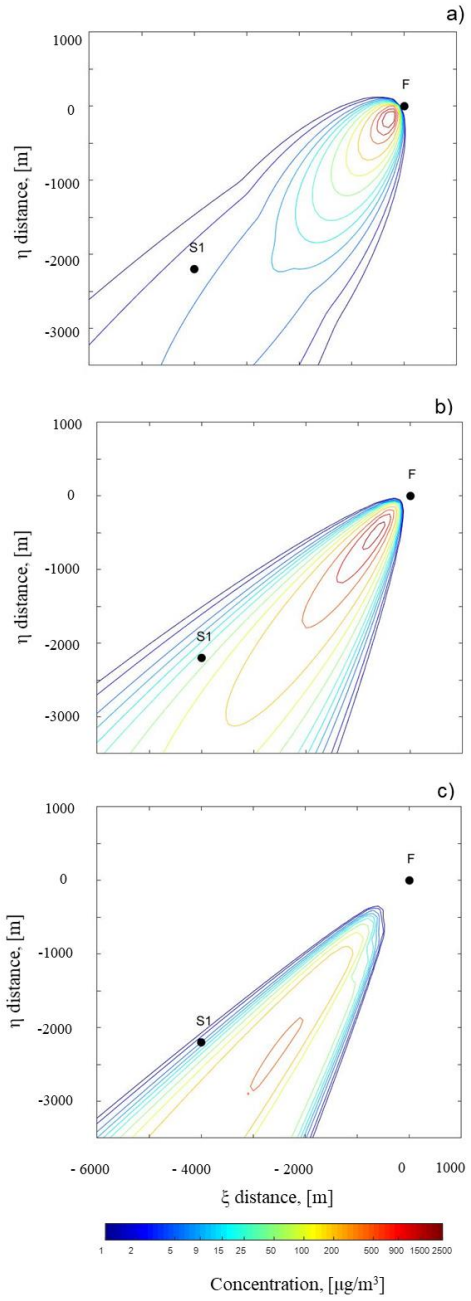


Figure III.4.4. *PM2.5 plume dispersion in prevailing wind conditions at 7 m/s for: a) very unstable (A), b) slightly unstable (C), c) slightly stable (E) wind field.*

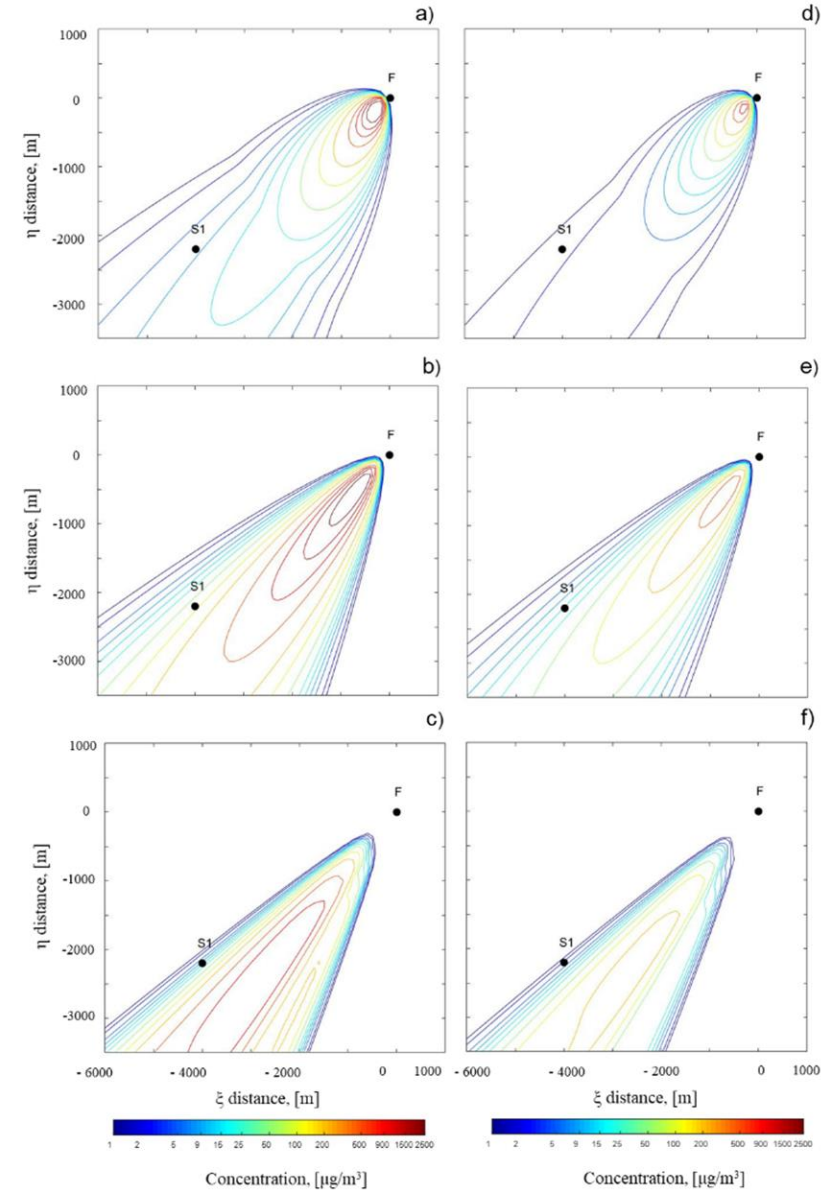


Figure III.4.5. *PM2.5 plume dispersion in prevailing wind conditions at different wind intensity and stability classes: 3 m/s wind intensity for a) very unstable (A), b) slightly unstable (C), c) slightly stable (E) wind field; 15 m/s wind intensity for d) very unstable (A), e) slightly unstable (C), f) slightly stable (E) wind field.*

Conclusions

A high time resolution monitoring station, located 4 km from a fire occurring in the industrial area of Avellino (Italy) on 13/09/2019, and 500 m aside from the downwind direction, collected air pollution data during the event. Results indicated that the fire did not negatively affect the air quality of the city centre. In fact, the daily average values of PM2.5 were never above the law limit. The fire was extinguished after many hours of work and its effect in the city centre were observed after a few hours from the fire beginning, with the measured concentration of PM2.5 in the range between 12 and 41 $\mu g/m^3$. A single stack Gaussian Plume model approach was applied and validated with the available measurements to calculate the PM2.5 concentrations on the ground in a wide area downwind the fire location. Model results indicated that much higher ground level concentrations were produced during the fire, as high as 500 $\mu g/m^3$. Different wind and air stability conditions could have worsened the fire effects. In particular, lower wind speed and greater air stability could have induced a potentially more significant danger for the population living in the area. This result highlight that a few pollutant measurement stations are insufficient to properly monitor an area or to provide validation data for any predicting model. For a better protection of the population exposed, installation of a denser network of high time resolution monitoring devices would allow a timely and a reliable report on the ground level for any fire position.

III.5. Effect of the first lockdown due Covid-19 on the Milan air quality

Like the rest of the world, starting with the identification of the first case of Sars-Cov2 on January 31, 2020, Italy was also hit by the pandemic wave with the consequent lockdown of all activities. During this period, which took place between February and April 2020, the world has become an open-air laboratory in which the effects on air quality are due to the reduction of most of the polluting anthropogenic activities. Italy is characterized by a complex orography and different climatic zones can be identified, where, depending on the seasonal period, very favorable conditions for the pollutants accumulation and formation in the atmosphere can occur. A striking example is the Po valley and some areas of central-southern Lazio where during winter pollutants accumulate and coastal areas that have conditions that generally favor dispersion and reduce the possibility of secondary pollutants forming. The period of March is usually less favorable to the accumulation of pollutants than the months of January and February, when conditions of thermal inversion at low altitude and atmospheric stability often occur (REPORT SNPA 17/2020) with high values of the main pollutants, consequently the comparison of the levels observed during the lockdown period must be made, as a first approximation, with reference to the observations of the previous years in the same late winter-spring period.

From the report of the National Service for Environmental Protection (SNPA) 2020 it emerges that throughout the Italian peninsula there has been a reduction in the concentrations of nitrogen oxides (monoxide, which is the prevalent form emitted directly) and dioxide, (partly emitted directly and, partly formed in the atmosphere) of carbon monoxide, and benzene. This reduction is for nitrogen dioxide on average around 40%, which during the lockdown ranged from a few percentage points to values over 70% in some sites. If it was easy to define the trends for nitrogen oxides, the situation relating to particulate matter (PM10, PM2.5) is less linear. It should be remembered that particulate matter is a complex mixture of solid and liquid particles dispersed in the atmosphere that has three macro-components, which have different relevance depending on the season and the geographical area. Two components are of natural origin, the marine aerosol and the particles that originate from long-distance transport of desert sands, the third, terrigenous component or rising from the ground, partly has a natural origin and partly can bring back particles originally emitted into dispersion from vehicles due to tire brake, or produced by construction activities, shipyards or released by mining activities.

Particulate matter concentrations vary, not only with meteorology and primary PM emissions from anthropogenic sources but also with emissions

from natural sources, difficult to predict and highly variable from year to year, and precursor gas emissions from different sources. Hence, the behavior of emissions and PM formation during lockdown is complex. For example, in some regions, as people have been forced to stay indoors, there may have been an increase in primary PM emissions from domestic burning of coal or wood, while primary traffic PM emissions have been significantly reduced. Agricultural emissions of primary PM were probably not affected by the lockdown, while some industrial emissions (e.g. PM and nitrogen and sulfur oxides, NO_X and SO_X) were reduced in various sites and countries. (EEA 2020). Certainly during the lockdown period, very significant reductions in traffic flows were observed which reached, on a national basis, starting from the middle of March about 70% for light vehicles and 38% for heavy vehicles, for then gradually return to previous levels in the first half of June. This decrease certainly influenced the decrease in particulate levels which, however, were, in some cases, offset by emissions from domestic heating. In this regard, no particular reductions have been observed, indeed from preliminary estimates in some areas, also due to a month of March on average colder than usual, it seems possible that there was also in some areas a moderate increase in emissions compared to the average. of the period (ARPA VENETO). Furthermore, it is necessary to take into account the contributions related to industrial and livestock activities and transport due to Saharan dust which in some cases can be added to the particulate already produced in the urban environment. All of these considerations may constitute a preliminary motivation for the minor, and insignificant in some cases, reduction in the levels of PM10 and PM2.5 recorded during the lockdown period in Italy.

In particular, the problem of atmospheric pollution has always been a peculiarity of northern Italy due to the large concentration of industrial activities. In this part of the thesis work, the effect of the first lockdown was analyzed, comparing the concentrations of the different pollutants in the different two scenarios, pre and post lockdown throughout the Po Valley. More in detail, the comparison was made on the metropolitan city of Milan (Italy), as it is particularly representative, due to the characteristics of the urban fabric and variety of the industrial factories. The metropolitan city of Milan is one of the largest and most populous in northern Italy with approximately of 1,400,000 inhabitants in the municipality alone and more than 3 million inhabitants in the metropolitan area. It is characterized by the many activities of different nature by which it is daily interested. In particular, its industrial poles, the massive urbanization and the orography not favorable to the dispersion of pollutants, make it, together with Turin, Venice, Naples and Cagliari, among the cities where the greatest overruns of average concentrations are recorded. daily of PM10 (Rapporto mobilit'aria 2019). The mountains surround the valley to the north and west (Alps) and south (Apennines), while the east side is open to the Adriatic Sea. This

topography favors air stagnation events and is accompanied by frequent winter thermal inversions and diffuse fog events that lead to the accumulation of particulate (PM) pollution (Carbone et al., 2010; Pernigotti et al., 2012; Larsen et al., 2012; Squizzato et al., 2013; Pecorari et al., 2013; Perrino et al., 2014).

Data source

The air quality data used in this part of the study are those of the one provided by the ARPALombardia network in Milan. ARPA Lombardias' air quality monitoring network consists of 85 fixed stations on the tregional territory of which, by means of automatic analyzers, continuously provide data at regular time intervals (generally on an hourly basis). The pollutant species monitored continuously are NO_X, SO_2, CO, O_3, PM10, PM2.5 and benzene. Depending on the environmental context (urban, industrial, traffic, rural, etc.) in which the monitoring station is active, the type of pollutants that must be detected is different. Therefore, not all stations are equipped with the same analytical instrumentation. The regional stations are distributed throughout the region according to the population density and the type of territory, respecting the criteria defined by Legislative Decree 155/2010. The data provided by the fixed stations are integrated with those collected during temporary measurement campaigns through mobile laboratories. In particular, the data used in the case study are the ones measured in Milan where he network is composed by 5 monitoring station that provides for the PM10 measurements and only 2 of them provides also for the PM2.5 measurements.

Results

To verify whether there has actually been an improvement in air quality, it is necessary to start from the analysis of trends relating to PM10 concentrations over the last 10 years. Figure III.5.1 shows the annual averages recorded by the regional air quality network over the last ten years and it can be seen that the annual average limit of 40 $\mu g/m^3$ (according to D.Lgs 155/2010) has not been exceeded in the years 2010, 2013-2014, 2016 with a decreasing trend between 2017 and 2020.

The trend of pollutants aero-dispersed is strongly linked to the meteorological variables that characterize each year. Therefore, to better visualize the multi-year trend, eliminating the variability between contiguous years due to different meteorology, the five-year moving average was calculated. In this way, considering that the meteorological difference over five-year periods is less evident, the trend linked to the trend of emissions is highlighted more clearly. Figure III.5.2 and Table III.5.1 show the average of the first six months of 2020 to get a first pollution level evaluation of the

first half of the year. The analysis of the 5-year moving average shows a progressive decrease in the average levels of pollution in the years 2010-2020. This decrease can be associated with the implementation of more stricted environmental policies and the improvement of technology in the emissions abatement and reduction.

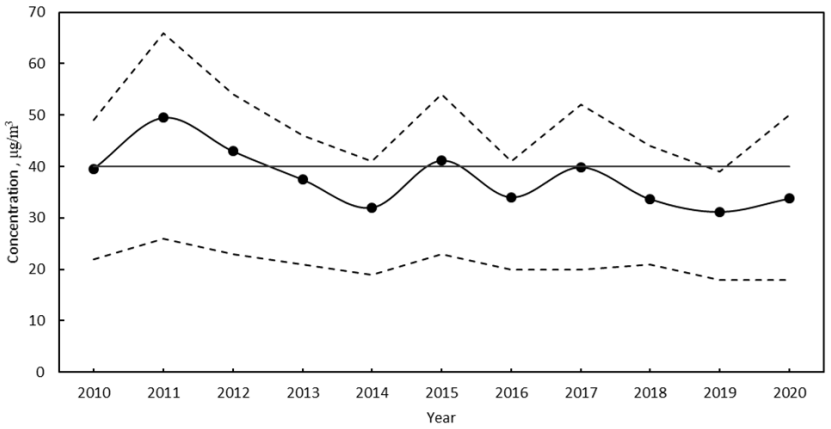


Figure III.5.1. *PM10 annual average concentrations in the period 2010-2020 measured by the ARPALombardia air quality network in Milan (black solid line). 25th and 75th percentile of annual average concentrations (dashed black line). PM10 annual law limit of 40 µg/m^3 according D.Lgs. 155/2010 (black constant line).*

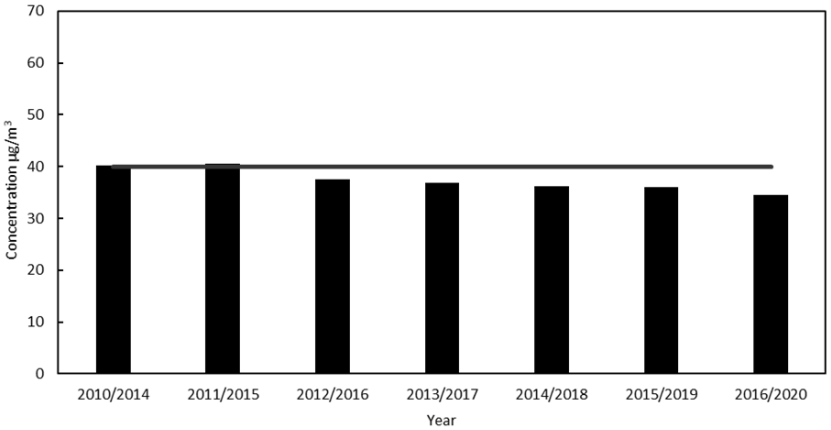


Figure III.5.2. *PM10 5- year mobile annual average concentrations measured in Milan by the ARPALombardia air quality network. PM10 annual law limit of 40 µg/m^3 according D.Lgs. 155/2010 (black constant line).*

Table III.5.1. *PM10 5- year mobile annual average concentrations in the period 2010-2020 measured by the ARPALombardia air quality network in Milan.*

Year	5- year mobile annual average concentrations μg/m^3
2010/2014	40
2011/2015	41
2012/2016	38
2013/2017	37
2014/2018	36
2015/2019	36
2016/2020	34

To analyze the pollution trend in the city of Milan during the lockdown, it is useful to divide the time series into two periods. The first between 23/02/2020 and 8/03/2020 in which the restrictive measures gradually began and from 9/03/2020 to 4/05/2020 in which the measures are containment and the lockdown was more drastic. It can be seen from Figure III.5.3 that up to 9/03/2020 the pollution levels are low. There would seem to be an increase in pollution on 17-20/03/2020 days. The growing curve in these days represents an accumulation of pollutants due to the phenomena of high pressure and low wind intensity recorded. In fact, these atmospheric phenomena favor stagnation and the accumulation of pollutants. Moving forward in the month of March, a very high peak around 90 μg/m^3 is noted on the days 28-29/03/2020. If we refer to the ratio between the concentrations of PM2.5 and PM10 (black curve in Figure III.5.3), the low value of this ratio on days 28-29/03 indicates a strong external influence on pollution. According to the global aerosol "Copernicus Atmosphere Monitoring Service" forecast, in those days a mass of dust from Central Asia hit Europe, contributes to raise the measured particulate levels. The transported particulate is of desert and sandy origin mainly composed of siliceous materials unlike PM10 of urban origin characterized by the presence of carbon and metal based particles.

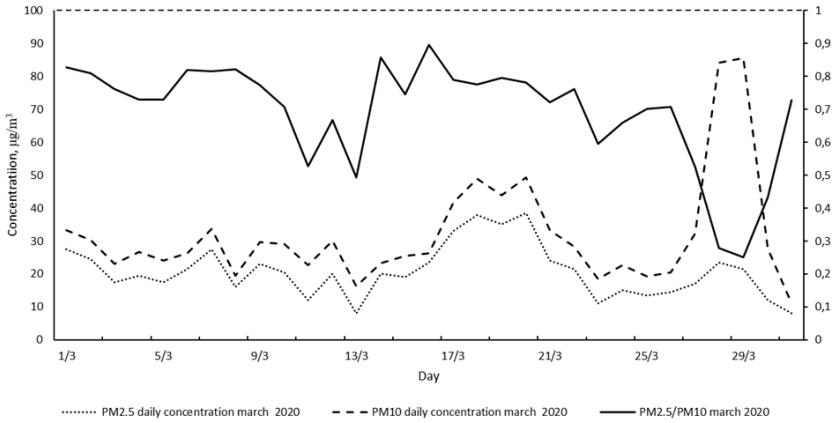


Figure III.5.3. *PM10 daily average concentration (dashed black line), PM2.5 daily average concentrations (dotted black line), PM2.5 and PM10 concentrations ratio (solid black line), measured by the ARPALombardia air quality network in Milan during march 2020.*

From a climatic point of view, the Mediterranean atmosphere (called MED) is characterized by rainy winter seasons and hot and dry summer that affect the continental areas that surround it. It is therefore full of particulate of anthropogenic origin, coming from the vast European industrialized regions, and of natural particulate of crustal origin, coming from the extensive arid and semi-arid areas of Africa and the Middle East, and of aerosols of marine origin, generated by the surface of the MED, that of volcanic dust, emitted by the main volcanoes of the basin, such as Etna and Stromboli (Arnold et al. 1982).

There are some methodologies in the literature to estimate the contribution of Saharan dust on urban pollution. The analysis by Bonasoni et al. (Bonasoni P. et al., 2004) confirms that the contribution of the mineral aerosol on the urban PM10 values can be very critical, favoring the threshold of exceedances. The study shows that the average increase in PM10 (ΔPM10) due to Saharan dust in the Po Valley area is within a range of 20-40% compared to the typical seasonal values. Considering the mean difference of PM1. And PM2.5 concentrations measured in March 2020 it was possible to evaluate an average ΔPM10. Adding the value of ΔPM10 calculated to the PM2.5 values measured in 28-29/03/2020, it was possible to obtain the PM10 concentrations without the contribution of Saharan dust and that only takes into account only the anthropogenic source.

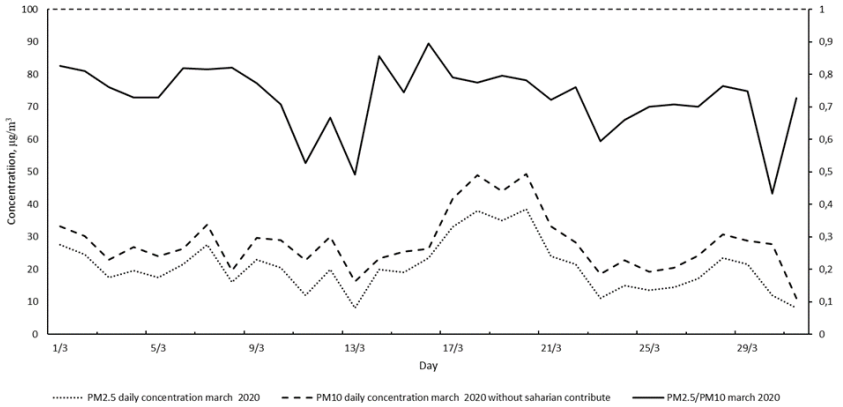


Figure III.5.4. *PM10 daily average concentration without the Saharian contribute (dashed black line), PM2.5 daily average concentrations (dotted black line), PM2.5 and PM10 concentrations ratio (solid black line), measured by the ARPALombardia air quality network in Milan during march 2020.*

As shown in Figure III.5.3, it can be seen that the contribution of Saharan dust to PM10 in the days 28-29 / 03/2020 is very high and that by removing their contribution from the measured values, the values obtained fall within the monthly trends obtained (Fig. III.5.4). Saharan dusts are not harmful to the human body as, given their nature, they do not produce harmful effects on human health unlike the particulates produced in an urban environment. Therefore, it is useful to separate their contribution from the total PM10 measured in order to have a clear estimate of the contribution of particulate matter due only to activities related to the urban context. After the annual pollution trends analysis of, it is useful to analyze in detail the differences in the 2010-2020 period, especially between the months of January, February, March and April that characterized the lockdown. From the values shown in Figure III.5.5 it can be seen that the month of March 2020, equal to 2016, shows the lowest concentration value of PM10. The cyclical trend of pollution suggests that the concentration of particulate matter measured in January is always greater than that measured in February. From Figure III.5.5 there is evidence that in some years (2011, 2012, 2017) the concentrations measured in January are lower than in February. This phenomenon could be evidence of the permanence of particles in the atmosphere. In fact, it is known that both PM10 and PM2.5 particles can remain in the atmosphere for several weeks accumulating. The month of April 2020 appears to have low values, similar to those recorded in 2012 and 2019. The month of June 2020 is the one with the lowest concentration ever measured since 2010.

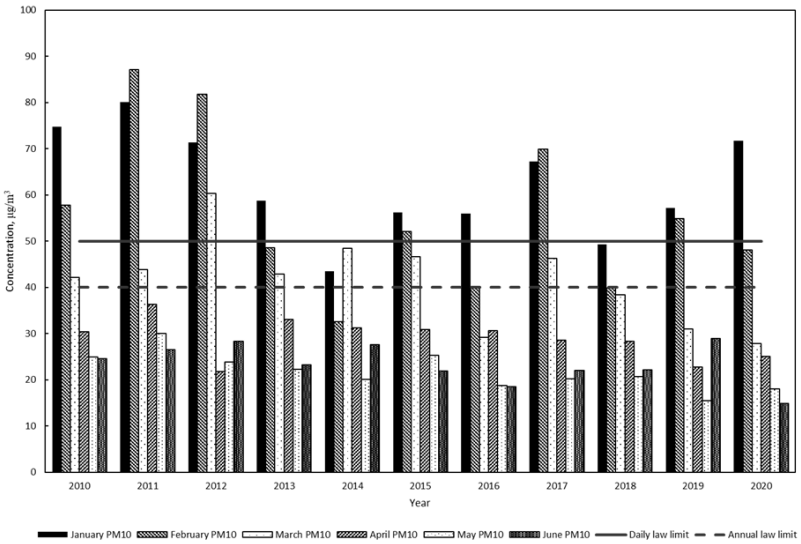


Figure III.5.5. *Monthly PM10 average concentrations for the first six months of 2020 measured in Milan by the ARPALombarrdia air quality network.PM10 monthly law limit of 50 µg/m^3(constant black line) and PM10 annual law limit of 40 µg/m^3 (constant dashed black line) according D.Lgs. 155/2010.*

The lockdown affected the months of March and April 2020. It is useful, therefore, to include the month of April in the analysis as well. In particular, the trend of PM10 fine particles in the months of March and April 2018-2020 was analyzed. Figure III.4.6 shows a gradual decrease in the average concentration of PM10 in March by 8.15% from 2019 to 2020 and by 16% compared to 2018. The month of April, compared with 2018 and 2019, shows a reduction in the 11% compared to 2018 but an increase of about 10% compared to 2019.

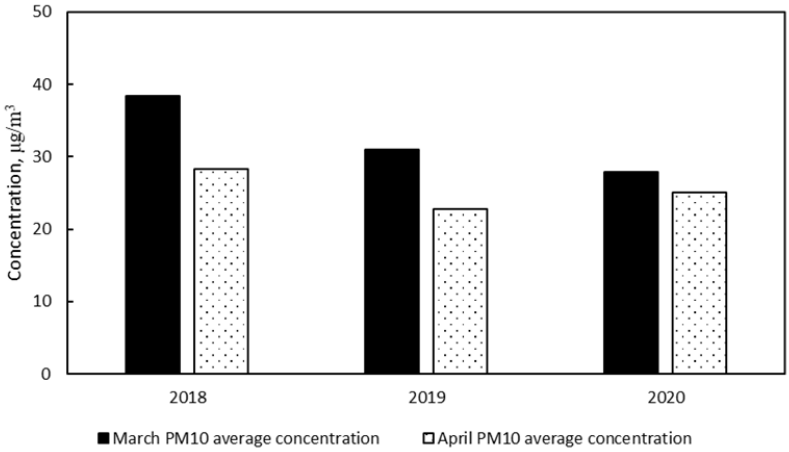


Figure III.5.6. *PM10 monthly average concentration during march and april in the period 2018-2020 2020 measured in Milan by the ARPALombarrdia air quality network.*

To analyze in more detail the variation of measured particulate matter, it is important to take into consideration the meteorological aspect that influences this phenomenon. In detail, for the months of March in the years 2010-2020, the average temperature variation recorded, the number of days and the cumulative millimetres of rain were considered. From the data shown in Figures III.5.7 and III.5.8 it can be seen that the month of March 2020 recorded a quantity of rain higher than 2019 but in line with the average of previous years, maintaining a number of rainy days similar to previous years, with the exception of 2018. which turns out to be the year in which March recorded the highest number of rainy days of the decade.

Although the meteorological values are similar to previous years, in March 2020 there was a decrease in the concentration of PM10 due to the contingent situation following the lockdown. The analysis of the average temperature is an indicator of the coldness of the winter which may have led to a higher use of domestic heating. Given the comparable average temperature between the months of March 2019 and 2020, it can reasonably be assumed that although the lockdown forced the citizens into their homes who therefore kept the domestic heaters on for longer, they did not contribute to the increase in levels of heating. PM10 which if compared with the month of March 2019 are lower. Therefore, the lower PM10 share measured in March 2020 compared to previous years is due to the reduction in vehicular traffic and the suspension of work activities due to the lockdown.

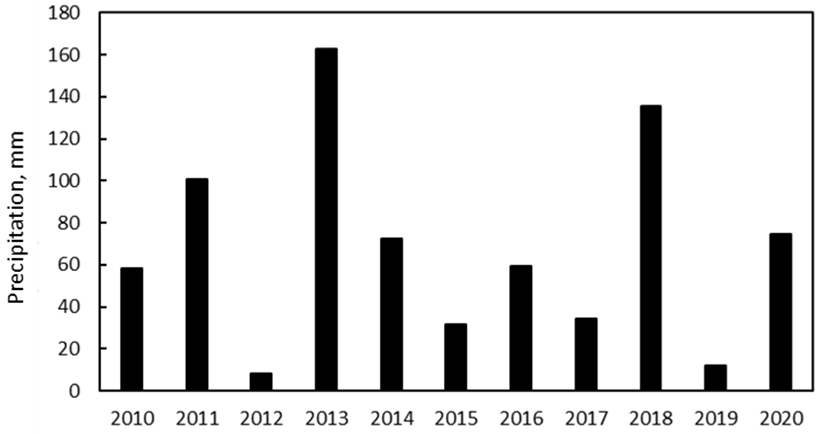


Figure III.5.7. *Rain cumulative daily value measured by the ARPALombardia meteorological network.*

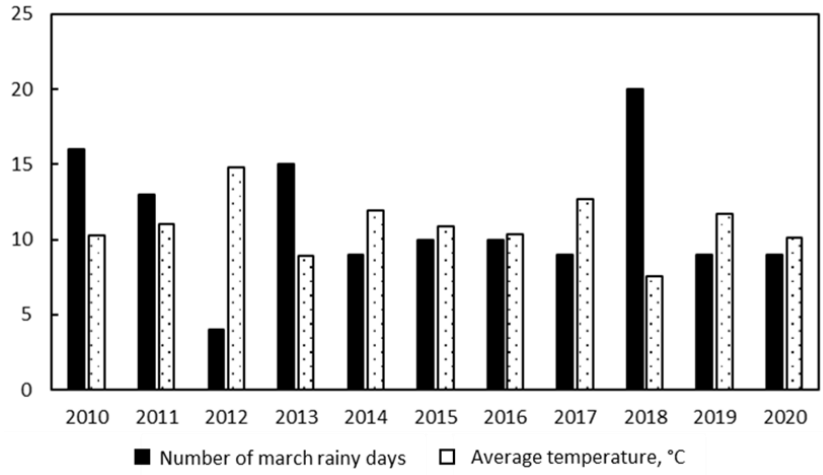


Figure III.5.8. *Rain cumulative daily value measured by the ARPALombardia meteorological network.*

After all this considerations, it is clear that during the last ten years, air quality in Milan improves. Moreover, during the first lockdown the reduction in air pollution is evident.

Conclusions

The influence of the lockdown on pollution in the Milan area is not trivial. In order to correctly understand how the current situation has affected pollution, it is necessary to investigate multiple factors. It was possible to verify, with some certainty the decreasing evolution of pollution levels in the

years 2010-2020, define the specific trends of PM10 for the years 2010-2020 in the months of January February and March, investigate the trends of pollutants only for the months of January February and March in the years 2018-2020. From all the data analysed was clear that the pollution reduction during 2020 was due to the limitation occurred during lockdown.

III.6. Validation of the Real-time on-road monitoring stations results

As an example of the collected results, the daily average concentrations of PM10, PM2.5 and, NO_2 in the city districts are reported directly on the map in Figure III.6.1. The mean concentration of PM10 (Fig. III.6.1b) is around 38 µg/m^3 for all the districts investigated and around 8 µg/m^3 and 15 µg/m^3 for the PM2.5 (Fig.III.6.1a) in district 3 and districts 2, all values below the law limit. The concentration of NO_2 (Fig. III.6.1c) is also low, and the average of the three districts is almost 14ppb. Each mean value concentration represents the average value of the data obtained from different vans passed through the same district during the day. In this way, it is possible to have a huge amount of information about the same district from different measuring devices with a traditionally fixed monitoring station is limited.

Figure III.6.1. *Map of a) PM2.5 mean concentration, b) PM10 mean concentration, c) NO_2 mean concentration divided by districts. Color codes are provided in Table 12. Red represents pollutant concentrations above the air quality limits, yellow indicates an almost good air quality and green indicates a good air quality. Grey represents the absence of data.*

Table III.6.1. *Color codes.*

	PM10 (µg/m^3)		PM2.5 (µg/m^3)		NO_2 (ppb)		VOC (ppb eq)	
Color	**min**	**max**	**min**	**max**	**min**	**max**	**min**	**max**
Green	0	20	0	14	0	99	0	10
Yellow	21	49	15	24	100	199	11	24
Red	50	999	25	999	200	800	25	100

To make the visualization more immediate, the data collected are reported on the map according to a traffic light code reported in Table III.6.1. Red indicates zones overcoming of the law limit for the specified parameter. Yellow indicates zones with an almost good air quality and green

indicates zones with a good air quality. The evaluation of the mean pollutant concentrations using the grid method (Fig. III.6.2b) gIIIes detailed information about pollutants and highlight criticalities that are hidden in the overall view of Figure III.6.2a. In particular, Figure III.6.2 shows the differences in results coming from different procedures in data aggregation relative to the concentrations of PM2.5. Data analysis reveals that in cell marked red local averages can exceed of 8.8 μg/m^3 the thresholds of 25 μg/m^3 of PM2.5 established by the Italian law. Instead, data of district averages hide these critical situations by providing values below 13.33 μg/m^3, well within the law limit.

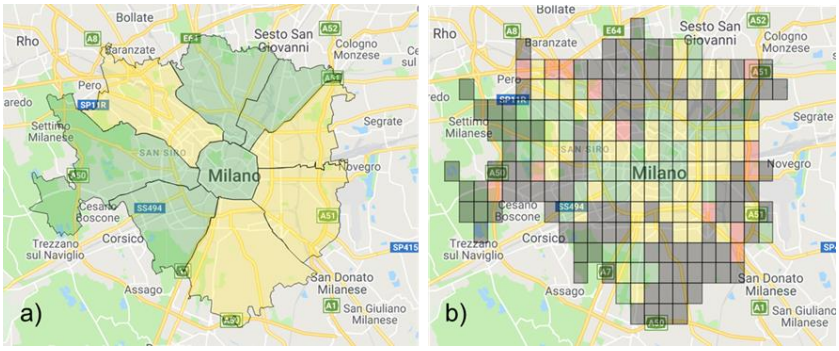


Figure III.6.2. *a) Map of PM2.5 concentrations divided by districts, b) Grid map of PM2.5 concentrations covering 1Km^2.*

In order to investigate the difference of pollution between working days and the weekends, two days were chosen as example. As it is shown in Figure III.6.3, during the weekends, the concentration of pollutants in some areas decrease. Data analysis indicates that the corresponding concentration of PM2.5 decreases of about 7% during the weekends (Fig. III.6.3b, f). In fact, the principal source of PM2.5 emissions is the daily traffic that during the weekends decreases. Data analysis indicate that during weekends, also the concentrations of PM10, NO_2 and VOC decrease, too but not as much to determine a change of color in the scale adopted (Fig.III.6.3).

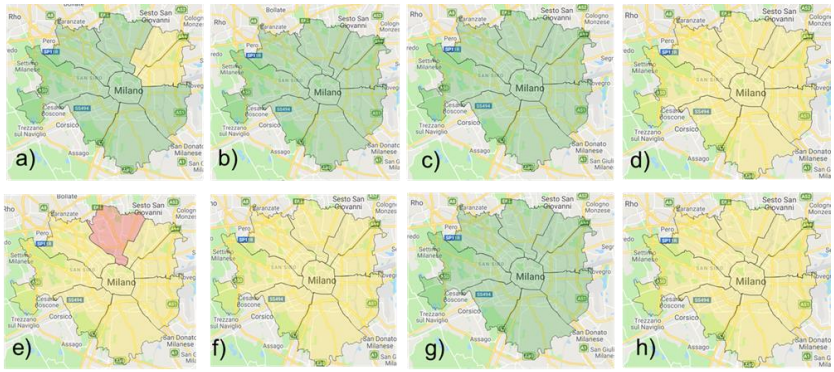


Figure III.6.3. *Map of pollutants concentration. For a working day (2018/08/23): a) PM10, b) PM2.5, c) NO_2, d) VOC. For the weekend (2018/09/28) e) PM10, f) PM2.5, g) NO_2, h) VOC.*

Comparison with ARPALombardia

To validate the data, some comparison with the city network has been done. The Regional Agency for the Environmental Protection of Lombardy (ARPALombardia) has a fixed monitoring network in Milan placed in via Senato (District 1), via Pascal (District 3) and viale Marche (District 9) that provides for the measurement of both PM2.5 and PM10 except for the S3 station. The data obtained from the ARPALombardia network are based on a different sampling method. In this case the measurement is by gravimetry as regulated in D. Lgs. 155/2010. The data are aggregated as daily averages and are available the day after the sampling due to the procedure of measurement.

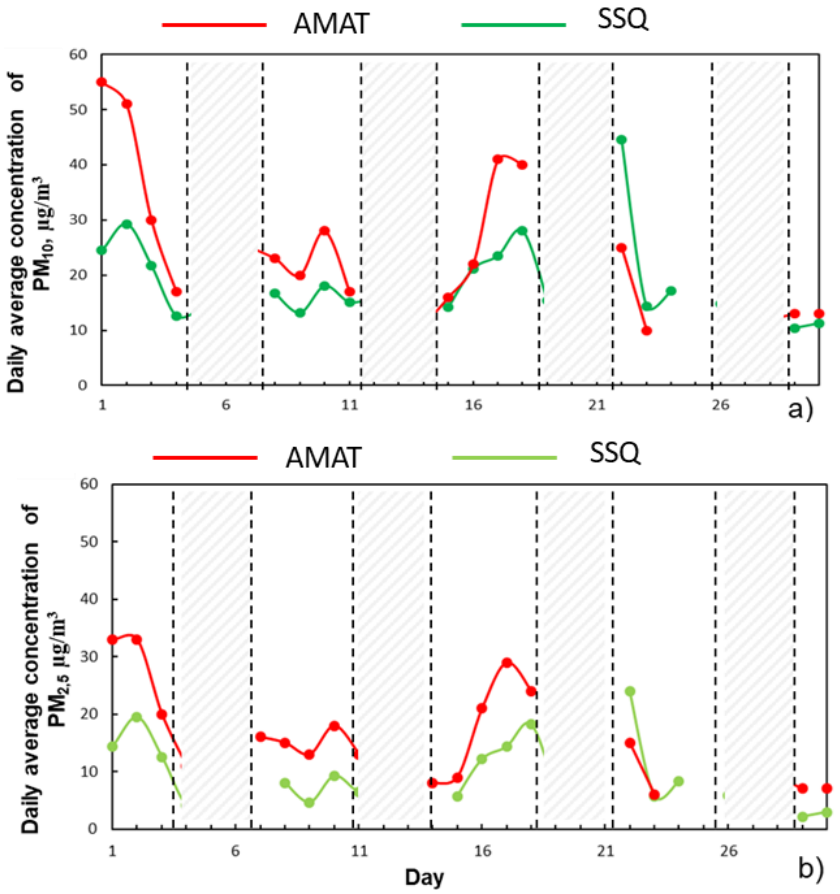


Figure III.6.4. *PM10 and PM2.5 daily average concentration for a) AMAT in via Senato measuring station and b) Sense Square ROMs for district 1 during 31 days.*

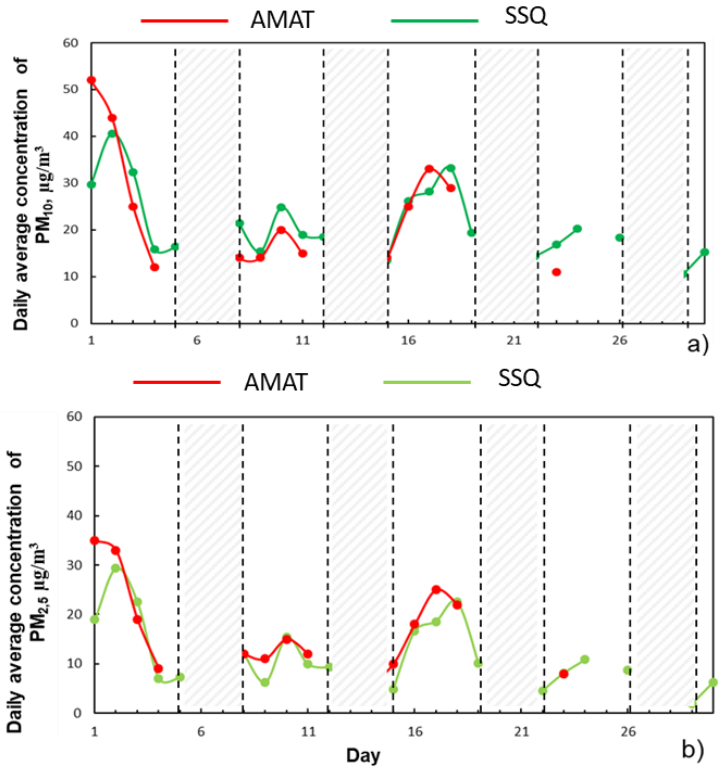


Figure III.6.5. *PM10 and PM2.5 daily average concentration for a) AMAT in via Pascal measuring station and b) Sense Square ROMs for district 3 during 31 days.*

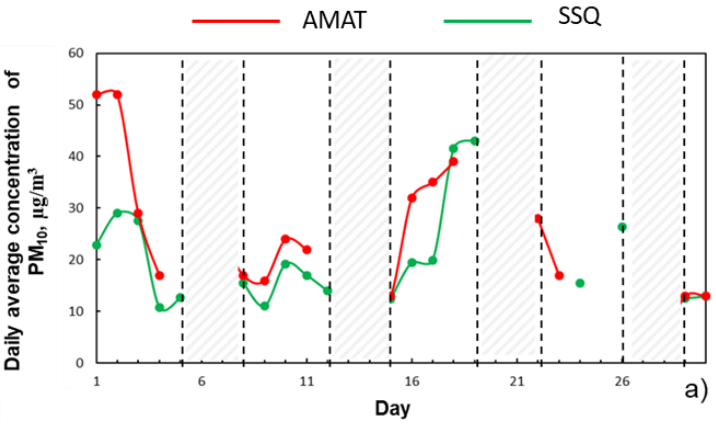


Figure III.6.6. *PM10 daily average concentration for a) AMAT in viale Marche measuring station and b) Sense Square ROMs for district 9 during 31 days.*

The value of concentrations of PM10 and PM2,5 measured by the ROMS are similar to the ones provided by the AMAT network (Fig. III.6.4, III.6.5, 5.6.7). The difference takes into account the different measurement used by the two air quality networks. The good accordance of the data makes the data provided by the ROMs robust and more affordable.

Cell network analysis

The almost 63600 data provided by the ROMs each day allows the detailed knowledge of the situation in the square cell. For each cell it is possible to obtain information about daily, weekly and hourly behaviours. The cell considered in this case belongs to District 5 (Fig. III.6.7) and is near the city centre of Milan. This particular cell is characterized by high anthropic activities and a high level of urbanization. During May in the cell can be highlighted four peaks of PM2.5 above the law limit. In particular, the one occurred the 8th of may represents an exceedance of PM2.5 concentration respect to the law limit of 25µg/m^3 while the other peaks occurred are higher than 75% of the law limit (Fig.III.6.8a). The PM10 concentrations recorded are low and above all they are far below the law limit of 50 µg/m^3 (Fig. III.6.7a). In this type of data representation (Fig. III.6.8), the red area represents an alarm range of PM2.5 values, the yellow area represents a warning set of values (always for PM2.5) and the green area represents a good air quality range of values for PM2.5.

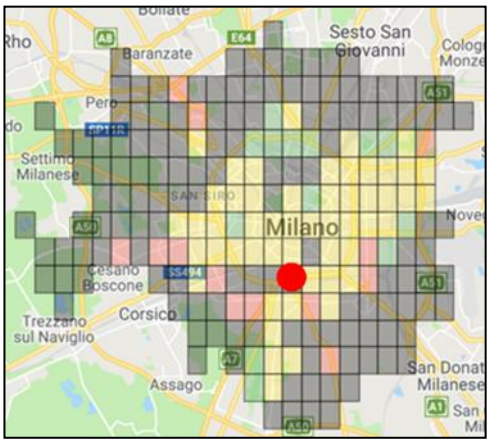


Figure III.6.7. *Square 5:2 in district 5.*

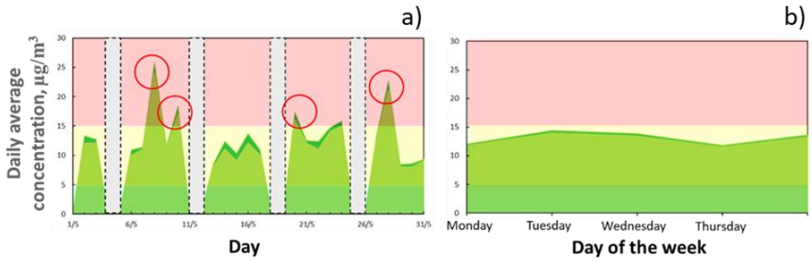


Figure III.6.8. *PM10 and PM25 daily average concentrations for the a) day of the months, b) day of the week in the cell 5:2. Higher values of concentrations are highlighted with a red circle. Green area represents the lowest concentrations, yellow area represents the alarm concentration and red are represent the exceedance concentrations.*

Considering the day of the week it is possible to note that in the cell analysed the day in which the levels of pollution are high are Tuesday, Wednesday and Friday (Fig. III.6.9b). The weekly behaviour of pollution depends above all on the anthropic activities that strongly influence for example the vehicular traffic. Tuesday and Wednesday represent the central days of the week in which the activities increase while Friday represents the day before the quiet. In fact, in most big cities, like Milan, Friday is the crucial day for departures; a lot of people use cars to move outside the city but at the same time a lot of tourists move towards the city.

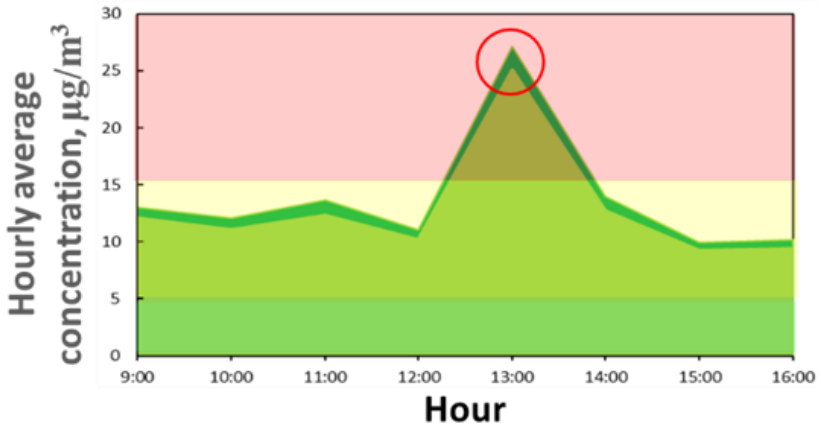


Figure III.6.9. *PM10 and PM25 hourly average concentrations for the ROMs operating range in the cell 5:2. Higher values of concentrations are highlighted with a red circle. Green area represents the lowest concentrations, yellow area represents the alarm concentration and red are represent the exceedance concentrations.*

Investigating the hourly behaviour of pollution in the operating range of the sensors a peak is registered around 13:00 (Fig. III.6.9). The high value of PM2.5 concentrations is probably due to the characteristics of the cell. In this part of the city there is a school and a lot of food activities. The time highlighted is typical of the end of lessons at school and most of people takes children by car and is the time when most people in offices moves for lunch.

Conclusions

The on-road network of monitoring stations (ROMs) for the air quality was set-up in order to provide high spatial resolution data in a wide area. As a result, the average value of the pollutant concentrations can be reported directly on the city map in order to have a clear definition of the environmental situation. Due to the quantity of data collected from the ROMs some information of the weekly pollution behavior can be elaborated. In particular, it is possible to see how the levels of the pollution change during the week. To provide a more detailed information the city map can be divided into grid of square cells. In each cell it is possible to calculate a representative value of the average concentration of each one of the pollutants taken into consideration.

III.7. Spatial interpolation of pollution data in a fine grid with missing data

III.7.1. Kriging methodology

Geostatistics studies the natural phenomena that develop on a spatial basis starting from the information deriving from their sampling. In particular, it studies the variability of the spatial parameters describing the aforementioned phenomena, extracting the rules in a reference modeling framework and using them to carry out the operations aimed at giving solutions to specific problems concerning the characterization and estimation of the phenomena investigated (Raspa and Bruno, 1994). In statistics, originally in geostatistics, kriging or Gaussian process regression is a method of interpolation for which the interpolated values are modelled by a Gaussian process governed by prior covariances. Under suitable assumptions on the priors, kriging gives the best linear unbiased prediction of the intermediate values. The method is widely used in the domain of spatial analysis and computer experiments. The reliability of a model is measured, in statistics by the standard error; in particular, with data dependent on each other the mean square error (MSE) is used.

The model

Consider a phenomenon that has spatial variability characteristics indicating with $z(\vec{s})$ the variable in the point of coordinates $\vec{s} = (u, v)$. The following definitions can be given:

- Regionalized variable (VR) is the function $z(\vec{s})$ which value depends on the localization and has a spatialized structure;
- Field (S): it is the domain where the VR can assume determined values and where the its variability it is studied.

The kriging interpolation consists in the prediction of an unknown value Z in the position s0, Z(s0), using a linear combination of the most reliable observations $\{Z(s_1), Z(s_2), Z(s_n)\}$, and fitting the linear combination coefficients by minimize the mean square error.
So, in general the method generates a predictor of the type:

$$p^*(Z; s_0) = \sum_{i=1}^{n} \lambda_i Z(s_i - m(s_i)) + m(s_0) \qquad \text{(III.7.1.1)}$$

where λ_i are the weight obtained minimizing the mean square error under the constraint of no distortion. If S the domain, according to the probabilistic approach, in each point $s \in S$ is defined a variate $Z(s)$ such that its mean m is defined as:

$$m(s) = \mathrm{E}[Z(s)] \qquad \text{s.t.} \qquad \mathrm{E}[Z] = \int_{-\infty}^{+\infty} z\, f(z)\, dz \qquad \text{(III.7.1.2)}$$

where E[Z] is the expected value of a continuous random variable with a probability density function $f(z)$. In fact, the assumption that the observations are distributed as a Gaussian sample, allows to have a conjugate Gaussian distribution (Z(s0), Z'), the average E(Z(s_0) | Z) linear in Z and dependent only on μ(s_i), and C(s_i, s_j) = cov(Z(s_i), Z(s_j)). Z' is a random variable that follow the same law p (conjugate) of Z. Since the best predictor able to minimize MSE is E(Z (s_0) | Z), in this case it is possible to obtain the linear predictor aforementioned.

Three types of kriging interpolation are distinguished in relation to the stationarity or otherwise of the considered process: simple (if stationary at given average), ordinary (if stationary with unknown average) and universal kriging (if not stationary, with averages expressed by a drift coefficient). In all three cases, the coefficients are determined by minimizing the mean square error under the non-distortion constraint:

$$\lambda = \underset{\lambda \in R^n | E[Z^*(s_0)] = [Z(s_0)]}{\arg min} Var[\textstyle\sum_{i=1}^{n} \lambda_i Z(s_i) - Z(s_0)] \qquad \text{(III.7.1.3)}$$

where, Z* represents a linear estimator. This translates mathematically into a linear system of Equations whose solution is exactly the set of coefficients {λ_i} that provide the best possible forecast (the one affected by the least uncertainty).

The coefficient (γ) of the matrix and the ones of the vector (λ) are evaluated starting from the variogram. γ represents the variogram of the variables considered. The variogram is a function $\gamma: \mathbb{R}^n \rightarrow [0, +\infty)$ that by receiving as input the distance between two points $\|h\| = \|s_i - s_j\|$ describes their spatial dependence on a certain stochastic process or to a specific variable field. The estimation of the variogram function is carried out on the basis of the data coming from the sampling of the phenomenon studied or from other indirect point data. The calculation of the experimental variogram γ^* (Fig.III.7.1.1) is carried out on the n points of the field of which information is obtained through an estimator:

$$\gamma^*(h) = \frac{1}{2n} \textstyle\sum_{i=1}^{n} [Z(s_i + h) - Z(xs_i)]^2 \qquad \text{(III.7.1.4)}$$

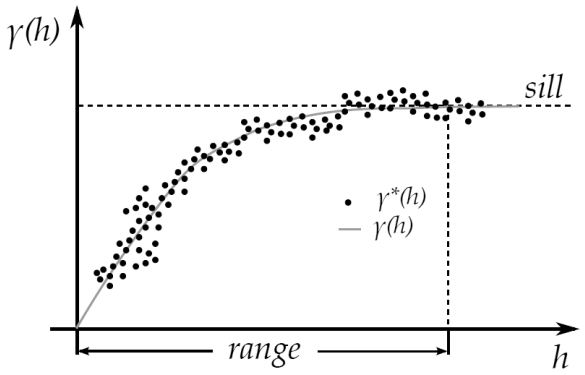


Figure III.7.1.1. *Experimental variogram with a sill fitted by the theory variogram.*

The trend of the experimental variogram as a function of *h* expresses the variability of the phenomenon considered with *h* and independently from the position, and reveals its stochastic behaviour, effectively suggesting the random function to be adopted. Given the lack of practicality of the experimental variograms, analytical functions are used which are commonly used to describe their behaviour. The geostatistical literature proposes some mathematical functions (Fig. III.7.1.1) suitable for describing the behaviour of the experimental variogram; once the appropriate function has been chosen, the theoretical variogram (γ) is obtained by modelling some parameters (usually with least squares techniques); the result is a continuous and derivable function $\forall h \in [0, +\infty)$.

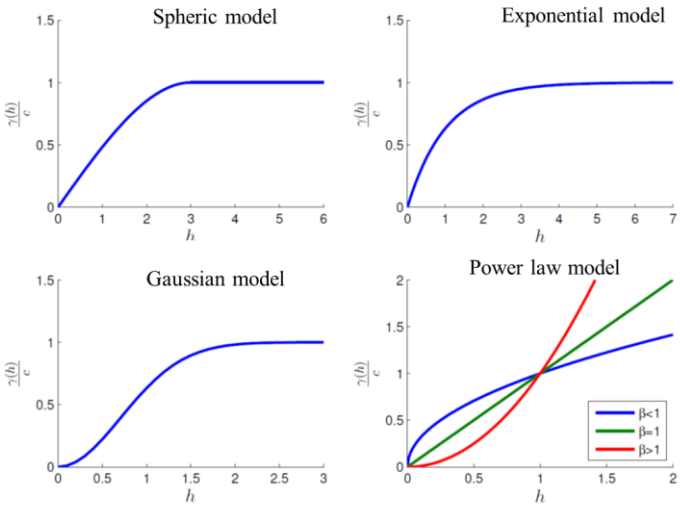


Figure III.7.1.2. *Qualitative graphs of basic theoretical variograms used in geostatistic.*

In the case of ordinary kriging, having an unknown mean μ, it is possible to build the model with the following assumption

$$Z(s) = \mu + \delta(s) \tag{III.7.1.5}$$

where $\{\delta(s): s \in S\}$ is a spatial white noise process with mean 0 and $var\{\delta(s)\} = \sigma^2$, consequently, the predictor will be:

$$p^*(Z; s_0) = \sum_{i=1}^{n} \lambda_i Z(s_i) \qquad \text{s.t} \qquad \sum_{i=1}^{n} \lambda_i = 1 \tag{III.7.1.6}$$

where the second Equation represents the non-distortion constraint. In this case it is necessary to solve the following linear system:

$$\begin{bmatrix} \gamma(0) & \gamma(\|s_1 - s_2\|) & \dots & \gamma(\|s_1 - s_n\|) & 1 \\ \gamma(\|s_2 - s_1\|) & \gamma(0) & \dots & \gamma(\|s_2 - s_n\|) & 1 \\ \vdots & & \ddots & & \vdots \\ \gamma(\|s_n - s_2\|) & \gamma(\|s_n - s_2\|) & \dots & \gamma(0) & 1 \\ 1 & 1 & \dots & 1 & 0 \end{bmatrix} \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_n \\ \nu \end{bmatrix} = \begin{bmatrix} \gamma(\|s_0 - s_1\|) \\ \gamma(\|s_0 - s_2\|) \\ \vdots \\ \gamma(\|s_0 - s_n\|) \\ 1 \end{bmatrix} \tag{III.7.1.7}$$

where the last row and the last column of the matrix are due to the relaxation by Lagrange multiplier (ν) of the non-distortion constraint. In this case the ordinary kriging, the variance can be expressed as:

$$\sigma_{OK}^2 = \sum_{1}^{n} \lambda_i \gamma_i(s_0 - s_i) + \nu \tag{III.7.1.8}$$

The pollutant dispersion in the air can be seen in a first approach as a stationary phenomenon, the ordinary kriging was implemented using the python package PiKrige. The code supports 2D and 3D ordinary and universal kriging. Standard variogram models (linear, power, spherical, Gaussian, exponential) are built in, but custom variogram models can also be used. In this case the code was run using a 2D ordinary kriging with a standard exponential variogram. Running the code using one-day data provided by the on-road monitoring network it was possible to predict the missing pollutant values.

Results

Appling the kriging function to the PM10 and PM2.5 concentrations to a one-day data, the grid in which the area is divided into is completed (Figs III.7.1.3b, e). The original grid (Figs. III.7.1.3a, d) has some missing values due to the not passage of the vans or temporary sensors malfunction. By applying ordinary Kriging, it was possible to estimate missing PM2.5 and PM10 concentrations. This first approach to spatial data estimation models, although very simplistic, allows a complete view of the city's environmental matrix. To better estimate the estimation coefficients, the model will be improved firstly considering the time-series of each square and matching the data on the concentrations of pollutants to significant variables for them. In particular, the vehicular traffic data characterizing each individual cell and the meteorological parameters which strongly influence the pollution trend will be used.

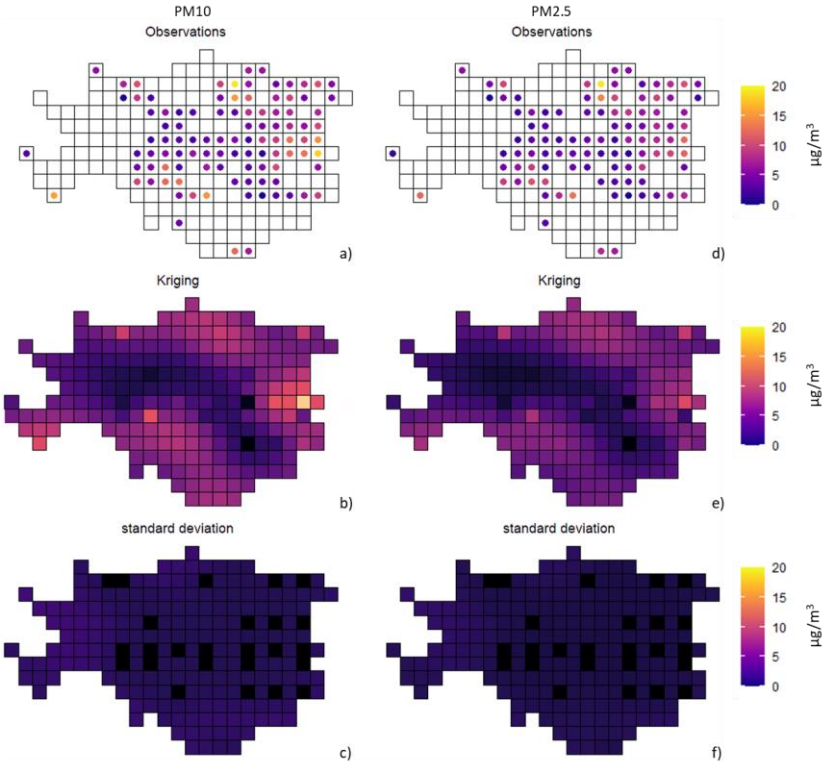


Figure III.7.1.3. *PM10 a) observed, b) estimated with kriging concentrations and c) standard deviation error. PM2.5 d) observed, e) estimated with kriging concentrations and f) standard deviation error.*

The values estimated with kriging (Figs. III.7.1.3b, d) show a good agreement with the observed data with lower values (blue and black) in the center of the mesh and higher at the margins, with values very close to the measured data. The standard errors are small, with maximum values corresponding to a few observed values (on the left and in the center of the map), however not higher than 2 (Figs. III.7.1.3c, f).

Conclusions

In order to have the overall view of the environmental situation of the city, the data obtained from the network of mobile stations have to be increased either by increasing the number of ROMs devices or with a combined use of the data given by ROMs and by traditional fixed monitoring stations. However, this last option requires a validation of the ROMs data compared with those of the fixed monitoring network installed in the city because, due to the different sampling methods adopted in the two cases. Ordinary Kriging model was applied to one-day data in order to fill all the grid cells with estimated day averages of pollutant concentrations.

III.7.2. Deterministic spatial interpolation

Raw concentration data, c, are transformed into deviation with respect to an assumed background value, c_{bg}, corresponding to the minimum concentration value recorded in the whole grid. Therefore, the interpolation procedure processes the deviation values $C=(c-c_{bg})$. In order to estimate missing concentration data, each cell, where measured data is available, can be considered as a pollutant emission source. As a result, a dispersion model with multiple sources is applied. Each measured data in a cell is the starting point of the spatial expansion provides its contribution to an expansion matrix. After the application of the expansion model to all measured values, the obtained matrices are combined to obtain a final matrix. The latter is given by linear combination of all the estimated matrices with a weight proportional to the distance from the emission source cell.

The spatial expansion takes into account the wind intensity and direction to model the pollutant dispersion phenomenology. The approach used in this study assumes a pollutant dispersion field such that each emission source could generate iso-concentration elliptical lines, with the source cell centre located in one of the focal points. The concentration in the points of each ellipse is equal to the measured concentration in the source cell multiplied by a decay factor depending on the ellipse size. It is assumed that the main axis of each ellipse is placed along the wind direction and the ratio between the ellipse axes is a function of the wind speed. Accordingly, the axes have been defined by the following Equations (Eq.III.7.2):

$$\begin{cases} a^2 + b^2 = 2r^2 \\ a^2 = b^2 + \gamma \mathrm{r}^2 v^2 \end{cases} \quad \text{(III.7.2.1)}$$

where a is the semi-major axis, b is the semi-minor axis, v is the normalized wind speed, r is the distance at which a given concentration decay coefficient is obtained in the absence of wind.

The estimated concentration in each cell in (*i*, *j*) position, $CE_{i,j}$, is a linear combination of the concentrations of the neighbouring source cells, $C_{k,l}$, properly weighted by means of factors corresponding to the source cells ellipses, passing through the (*i*, *j*) cell itself . In particular, the considered source cells are the closest eight on the grid placed around the (*k*, *l*) cell. The resulting concentration is gIIIen by the Eqs. (III.7.2.2) and (III.7.2.3):

$$CE_{i,j} = \sum_{\substack{k=i-1,\mathrm{i}+1 \\ k \neq i}} \sum_{\substack{l=j-1,j+1 \\ l \neq j}} C_{k,l} w_{k,l} \quad \text{(III.7.2.2)}$$

$$w_{k,l} = \frac{\alpha_{k,l}}{\sum_{\substack{k=i-1,\mathrm{i}+1 \\ k \neq i}} \sum_{\substack{l=j-1,j+1 \\ l \neq j}} \alpha_{k,l}} \Rightarrow \sum_{\substack{k=i-1,\mathrm{i}+1 \\ k \neq i}} \sum_{\substack{l=j-1,j+1 \\ l \neq j}} w_{k,l} = 1 \quad \text{(III.7.2.3)}$$

where $w_{k,l}$ is the weighing factor of the concentration contribution for the neighbouring cell in the (*k*, *l*) position and $\alpha_{k,l}$ is the corresponding decay coefficient. The decay coefficient was evaluated in two alternative ways. In the first case, $\alpha_{k,l}$ decreases linearly with the size $r_{k,l}$ of the ellipse, having a focus in the point in the (*k*, *l*) position and passing through the point in the (*i*, *j*) position, and defined by Eq (III.7.2.4):

$$\alpha_{k,l} = 1 - r_{k,l} d_c \quad \text{(III.7.2.4)}$$

where d_c.is a decay constant with d_c=0.05. In the second case, $\alpha_{k,l}$ is evaluated by an exponential decay function with d_c=1.57:

$$\alpha_{k,l} = \exp(-r_{k,l} d_c) \quad \text{(III.7.2.5)}$$

The parameter γ appearing in Eq. (III.7.2.1) is obtained by minimizing the deviation function, *f*, obtained as the sum of the squares of the residuals between the estimated concentration $CE_{i,j}$ and the measured concentration $C_{i,j}$, for the neighbouring cells where measured data are available:

$$f = \sum_k \sum_l \sqrt{(CE_{l,j} - C_{k,l})^2} \quad \text{(III.7.2.6)}$$

To evaluate the performance of the complete model, the normalized root mean square error between the measured values and the corresponding values estimated with the model was calculated as follows:

$$NRMSE = \sqrt{\frac{1}{N}\sum_k \sum_l \left(\frac{CE_{l,j} - C_{k,l}}{C_{k,l}}\right)^2} \qquad \text{(III.7.2.7)}$$

where N is the number of the values estimated.

Results

Preliminary calculations aimed at evaluating the optimal parameters and Equations of the spatial dispersion model. Figure III.7.2.1b shows that the f function, defined by Eq. (III.7.2.6), is weakly dependent on the parameter γ, probably due to the normalization of Eq. (III.7.2.3). Nevertheless, a minimum value can be observed and, thus, an optimal value for γ was set to 1.5. The parity plot of Figure III.7.2.1a compares the concentration values, estimated by Eq. (III.7.2.2), using for the decay coefficient $\alpha_{k,l}$ Eq. (III.7.2.4) or Eq. (III.7.2.5), respectively. Inspection of the Figure reveals that there is not a significant difference between the estimated values. As a result, the exponential decay function of Eq. (III.7.2.5) was adopted for the rest of the study since it yields intrinsically $\alpha_{k,l}$ values between 0 and 1.

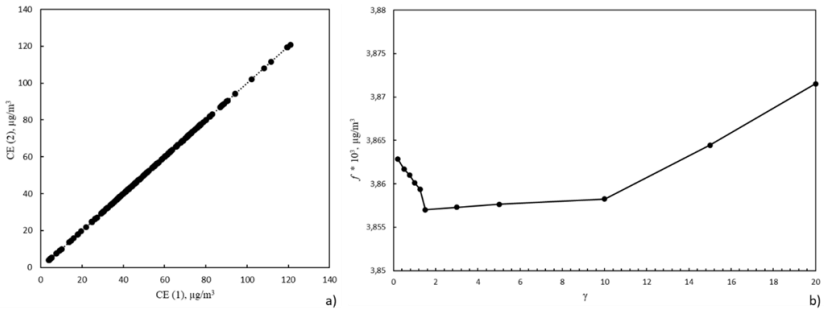


Figure III.7.2.1. *a) Comparison between CE concentration values estimated by Eq.(III.6.2.2): CE (1) with $\alpha_{k,l}$ according to Eq.(III.6.2.4), CE (2) with $\alpha_{k,l}$ according to Eq.(III.6.2.5). 40b) Minimization of f as a function of parameter γ.*

The grid (x, y) covering the investigated area of Figure III.7.2.2a, was rotated by an angle β corresponding to the wind direction. Figure III.7.2.3 reports the eight ellipses with constant decay coefficient values, calculated according to Eq. (III.7.2.5), passing through a selected (i, j) point, to estimate $CE_{i,j}$ as a function of the measured concentration in the eight

neighbouring cells. Inspection of the four panes, corresponding to different normalized wind intensity values, suggests that the higher is v , the longer is the ellipse major axis and the more effective is the pollutant dispersion. This scenario corresponds to lower $\alpha_{k,l}$ values for the source cells located downwind with respect to the (*i*, *j*) point and to higher $\alpha_{k,l}$ values for the source cells located upwind. As a result, the concentration $CE_{i,j}$ is affected by the concentration of the neighbouring cells with different weights. On the contrary, the lower is v, the shorter is the ellipse major axis and the less effective is the pollutant dispersion. In absence of wind (Fig. III.7.2.4), i.e. $v = 0$, the ellipses turn into circles corresponding to the lowest dispersion and the pollutant concentration tends to accumulate in the vicinity of the source cells. In this case, comparable $\alpha_{k,l}$ values are obtained for all the neighbouring cells and their corresponding concentration values affect the concentration $CE_{i,j}$ with the same weight.

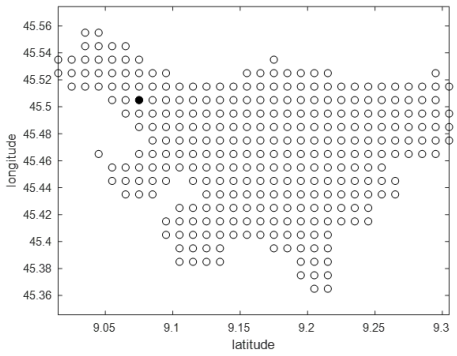


Figure III.7.2.2. *Example of point location on the measurement position map.*

Figure III.7.2.4 shows an example of the resulting PM10 and PM2.5 maps after the spatial interpolation according to Eq. (III.7.2.3) and using the measured concentration data, the wind intensity and direction. Concentration is represented according to a colour scale. Solid circles correspond to measured data, while hollow circles correspond to estimated data. The estimated values appear in good agreement with the close measured data. The estimate is less satisfactory in the periphery of the map where less measured points were available and the background concentration value c_{bg} was assumed. On the whole, the interpolation model appears effective considering that the error function *NRMSE* values, obtained according to Eq.

(III.7.27), were in the range of 0.08-0.7 for the majority of tested concentration maps.

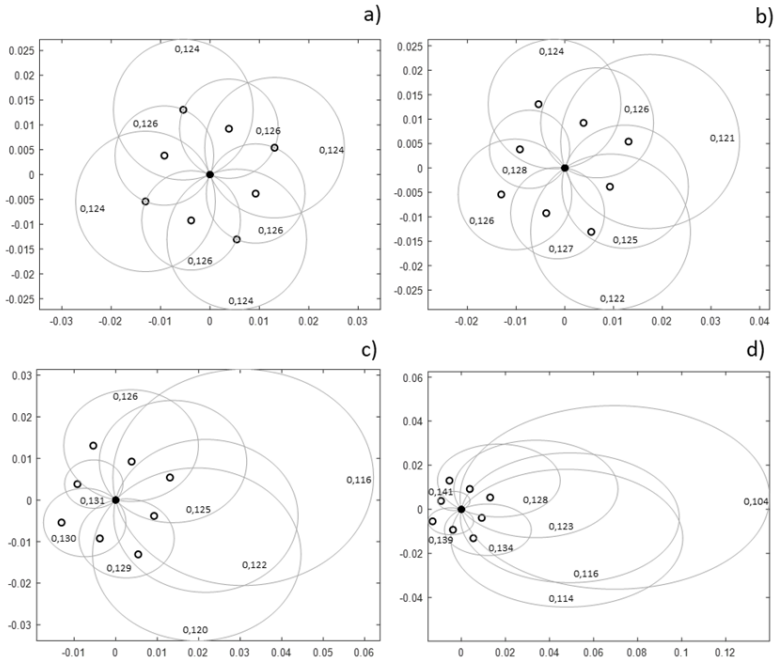


Figure III.7.2.3. *Decay coefficient elliptical fields to evaluate the concentration in the (i, j) point corresponding to the solid circle as a function of the concentrations measured in the eight neighbouring (k, l) points represented by the open circles for a)* v*=0, b)* v*=0.2, c)* v*=0.5, d)* v*=0.8 at wind direction of* 247.5°*. Corresponding* $\alpha_{k,l}$*values are reported on the curves.*

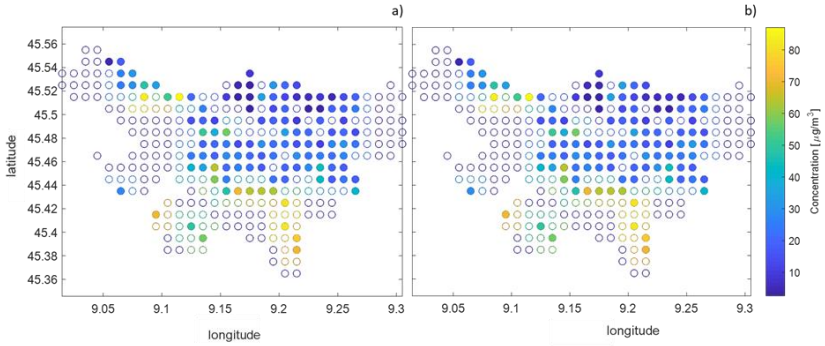


Figure III.7.2.4. *a) PM10 and b) PM2.5 concentrations estimated applying the spatial interpolation model. Solid circles correspond to measured data while hollow circles correspond to estimated data.*

Conclusions

Every day the monitoring sensor networks collects data about the air quality along the Italian territory. With this huge amount of data, it will be possible to expand the analysis done until now. The geo-spatialization techniques will be applied in more detail using appropriate algorithms for large datasets that allows the correct knowledge of the estimated concentration. The interpolation of missing values will be carried out on the ROMs data to evaluate the concentrations value during the not operating range of the sensors and they will be compared with the ones provided by the fixed monitoring stations that will be installed. The proposed interpolation method appears to describe qualitatively the effect of wind direction and intensity on the pollutant dispersion. The parameter relating the wind speed and the decay coefficient function $\alpha_{k,l}$ was optimized. This parameter plays a more important role than the mathematical function type chosen for $\alpha_{k,l}$ on the interpolation performance, provided that the weighing factors $w_{k,l}$ of the neighbouring concentrations are normalized. The resulting deviation between measured values of PM10 and those estimated by the proposed method is generally lower than 4 µg/m^3 and only in rare cases it approaches 10 µg/m^3 with NRMSE of about 0.3. These satisfactory results provide some confidence in the modelling approach and encourage further developments. Future work will consider a possible increase of the number of points with measured data to estimate the concentration in peripherical points. Finally, this type of spatial expansion could be used also in combination with other dispersion approaches like the Gaussian Plume and Puff dispersion models.

III.8. Comparison of spatial interpolation techniques for innovative air quality measurements

Air quality data

The data used in this study includes all measurements made by the ROM network in Milan between January and June 2020 (Lotrecchiano et al., 2019). The ROM network was made of 53 measuring devices located on courier vans. The monitoring devices allowed the measurements of gas and suspended particulate matter. In this study the analysis will be limited to measures relative to PM10. The design of a monitoring network involves a preliminary phase of defining the installation positions. These positions can be defined with a modeling approach, using operational street pollution models (Lotrecchiano et al., 2019).

The starting point is the subdivision of the ground surface in cells with a shape able to cover the whole ground surface over which the pollution is monitored by means of fixed or mobile sensors (e.g. Triangles, squares, or exagons). Raw data associated with the time and position from which they are taken are used to calculate calculating hourly averages of the pollution within the cell $c_{i,j}$ for the cell i and the time j. Given a time j, and considering the random position of the ROMs the available data will not be able to provide complete sets of values $c_{i,j}$. Hence the need to apply spatial interpolation methods.

Methods definition

Three methods for spatial expansion of data were compared. The first method considered is called of isoelliptical expansion (ISOE) was firstly developed by Lotrecchiano et al., (2021) and assumes that a local variation of the concentration of pollutants with respect to the city average affects the neighborhood with changes in the local concentration in the same direction, and that the points on the ground affected by the same concentration changes are located along elliptic iso-concentration curves in which the major axis has the same direction of wind and that the ellipse elongation increases with the wind intensity. In this work, a modification to the ISOE method is proposed, to account for the local specific changes by considering the history of the pollution of each cell in a one-month long period (MISOE method).

The effectiveness of these two models is tested against another frequently adopted method, which is called of the Inverse Distance Weighted (IDW), in which the distance between point is the only parameter affecting correlation, irrespective of the present atmospheric conditions. In the following, more

details of the methods tested are reported. For the description of the methods, it is assumed that the ground surface is dived in equal shaped cells (i.e squares, exagons, triangles) able to cover the whole surface and that a single value if each pollutant concentration is associated to the cell, ideally applied at the cell centroid (Fig.III.8.1).

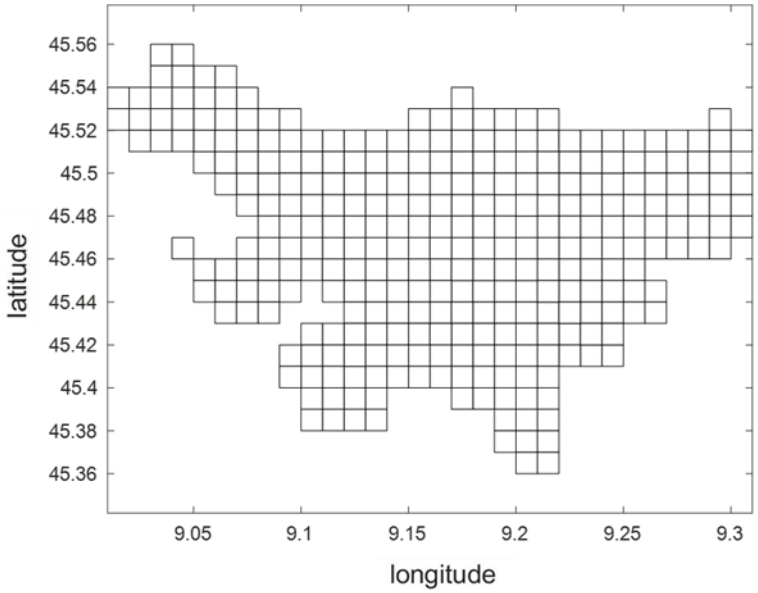


Figure III.8.1. *Square grid.*

ISOE model

The first step with the ISOE method is the calculation of the variation of local concertation at time j $c_{i,j}$, from the background level at the same time j, $c_{bg,j}$, that is

$$C = c - c_{bg} \tag{III.8.1}$$

The model considers that the cells interact with each other and the cell containing the measure can be assimilated to an emission source. The model developed takes into account the influence parameters relating to the pollutants dispersion and, in detail, to the wind intensity and direction. The model assumes each measured cell as an emission source and the pollutants disperse along iso-elliptical lines. The pollutants concentrations along the ellipses consider the measured concentration and a decay factor depending on the ellipse size. The ellipse lies on the wind direction and is function of the wind speed. The model considers that the major axis lies along the wind

direction and the relationship between the axes is a function of the wind speed. The axes of the ellipse are defined by the following system of Equations:

$$\begin{cases} a_j^2 + b_j^2 = 2r_{0j}^2 & \text{(III.8.2)} \\ a_j^2 = b_j^2 + \gamma r_{0j}^2 v_j^2 & \text{(III.8.3)} \end{cases}$$

were a_j and b_j are the semi-major and semi-minor axes length at time j, v_j is the wind speed normalized with respect to the historical maximum monthly measured value. time j, r_{0j} is the distance at which the decay along the ellipse is obtained for a wind speed equal to zero and γ is an adjustable model parameter relating the wind speed and the ellipse elongation. The parameter r_{0j} is associated to a decay factor α_j by means of a decay coefficient d_c:

$$\alpha_j = \exp\left(r_{0j} d_c \right) \quad \text{(III.8.4)}$$

Provided a second point that is placed at a distance r from the pollution source in a direction making an angle θ_j with the wind direction at time j, according to the ellipse Equation in polar coordinates, it is:

$$r = \frac{a_j\left(1-\lambda_j^2\right)}{1-\lambda_j \cos\theta_j} \quad \text{(III.8.5)}$$

Where the ellipse elongation, λ, is defined as follows:

$$\lambda_j = \frac{\sqrt{a_j^2 - b_j^2}}{a_j} \quad \text{(III.8.6)}$$

Combining (III.8.2) and (III.8.3), it is:

$$b_j^2 = a_j^2 \frac{2-\gamma v_j^2}{2+\gamma v_j^2} \quad \text{(III.8.7)}$$

Accordingly, Equation (III.8.6) becomes:

$$\lambda_j^2 = \frac{2\gamma v_j^2}{2+\gamma v_j^2} \quad \text{(III.8.8)}$$

Which allows to calculate a_j from Equation (III.8.6):

$$a_j = \frac{r(1-\lambda_j \cos\theta_j)}{1-\lambda_j^2} \quad \text{(III.8.9)}$$

So, using Equations III.8.9 and III.8.7 it is possible to calculate r_{0j} to be used in Equation (III.8.4) to calculate the local decay factor α_j

$$r_{0j}^2 = a_j^2 \frac{2}{2+\gamma v_j^2} \quad \text{(III.8.10)}$$

Considering the cell division of the ground, to each of which is associated a pollution level, in principle all the cells may affect the concentration of the others, with an effect that decreases with the distance according to a law that should at least reduce the effect by a factor comparable to the number of the other equally distant. With this in mind, it can be understood that the effect of the single cells not adjacent to the one considered would have only a second order effect and for this reason it was decided not to consider nonadjacent cells in the interpolation method adopted for the search of missing values. So, the estimated concentrations $CE_{i,j}$ in each target cell i at time j, are obtained as a linear combination of the neighbouring cells of the target cell. The set of neighbouring cells can be defined as follows: $i \in I$ where I is the set of cells; $K^{(i)}$ is the set of cell belonging to the neighbourhood of the target i t.c.; x_i and y_i are the Cartesian coordinates of the i target cells centroids. Furthermore;

$$K^{(i)} = \{k \in I: |x_k - x_i| \leq L_{xcell} \quad \wedge \quad |y_k - y_i| \leq L_{ycell} \wedge \ k \neq i\} \quad \text{(III.8.11)}$$

where L_{xcell} and L_{ycell} are the characteristic projections in the x and y direction of the cell size . In this case study, square cells parallel to the coordinate axes are considered and, therefore, $L_{xcell} = L_{ycell} = L_{cell}$ is the length of the cell side. $C_{k,j}^{(i)}$ are the observed concentration at time j in the cell k belonging to $K^{(i)}$. The estimated concentrations in the target cell $CE_{i,j}$ were calculated from a weighted combination of the observed concentration $C_{k,j}^{(i)}$.

$$CE_{i,j} = \sum_{k \in K} C_{k,j}^{(i)} w_{k,j}^{(i)} \quad \text{(III.8.12)}$$

where $w_{k,j}^{(i)}$ is the weighting factor for the concentration contribution of the k cell to the i cell at time j calculated as follows:

$$w_{k,j}^{(i)} = \frac{\alpha_{k,j}^{(i)}}{\sum_{k \in K} \alpha_{k,j}^{(i)}} \quad \text{(III.8.13)}$$

From Equation III.8.13 it appears that for equally distant neighbouring cells to the target cells, the coefficient d_c used in Equation III.8.4 is irrelevant. In our case with a square network, the centroid of the neighbouring cells along the diagonal are located a bit further from the centroid of target cells than the centroid of the cell on the sides. However, unless a strong decay coefficient, smaller than 1 km^{-1} is used, its value turned out to be scarcely affecting the final estimation value. For the calculations used in this paper, d_c was set to 1.57 km^{-1}. Finally, the estimated concentrations $ce_{i,j}$ are calculated from the inverse of Equation III.8.1:

$$ce_{i,j} = CE_{i,j} + c_{bg,j} \quad \text{(III.8.14)}$$

MISOE model

The MISOE model was obtained by modifying the previously developed ISOE model by introducing the local adjustment coefficient A_i^p to account for local variations of the measure which tend to keep the local measured value above or below the average and do not contribute to changing the neighbouring values. In fact, each cell is representative of different urban conditions and has a typical pollution level.

The use of local adjustment coefficient includes in the model the cell peculiarity. Since the availability of cell data depends on the specific day, the idea is that it might be possible to estimate these local adjustment coefficients considering a long enough time to have statistically representative values sequences but short enough to be still specific of the season considered. In order to calculate this coefficient, for each pollutant, it was considered the ratio between the local time monthly average, $M_{i,m}$, of the measured values $c_{i,j}$ values occurring in the cell i in that month, m, and the monthly average $M_{T,m}$ of all the $c_{i,j}$ values for the pollutant in all the cells within the town.

Namely, for each pollutant, for the position in the cell i, the month coefficient is calculated:

$$A_{i,m} = \frac{M_{i,m}}{M_{T,m}} \quad \text{(III.8.15)}$$

The descriptive coefficient of each cell $A_{i,m}$ is estimated month by month from the historical information available for each cell on a seasonal basis considering a number N_m of months. The global coefficient for each cell is obtained from the monthly coefficients calculated as follows:

$$A_{i,s} = \sum_m \frac{A_{i,m}}{N_m} \quad \text{(III.8.16)}$$

For the winter season, the months of November, December and January were considered and, therefore $N_m = 3$ while for the summer season, the months of April, May and June were considered (also $N_m = 3$).
So, considering the local adjustment coefficient, the estimated concentration calculated by Eq. 5.8.3 is modified as follows:

$$ce_{i,j} = A_i * \left(\sum_{k \in K} \frac{C_{k,j}^{(i)} + c_{bg,j}}{A_k} w_{k,j}^{(i)} \right) \quad \text{(III.8.17)}$$

Inverse distance weighted model

The inverse distance weighted model (IDW) is a widely used method which is based on the assumption (realistic for many phenomena) that the properties in a given point are more similar to those of nearby points, compared to those of more distant points (Bartier et al., 1996). Everything is related to everything else, but near values have a larger effect. For practical reasons a finite number N of cells in a finite neighborhood is considered in the calculation, according to Tobler's first law.

$$ce_{i,j} = \frac{\sum_{l=1}^{N} c_l r_{i,l}^{-2}}{\sum_{l=1}^{N} r_{i,l}^{-2}} \quad \text{(III.8.18)}$$

In this case the weights that will determine the value of the unknown position are inversely proportional to the square of the distance $r_{i,l}$ of the l - th point considered from the target cell i. The larger the distance, the less is the effect that points distant from the point where the concentration is unknown will have during the interpolation process. Although the assumptions underlying the IDW method are simple, in applying this method it is good to consider that:

- The power and the number of source points/radius to be interpolated is arbitrary.
- Estimates of uncertainty and possible errors are not carried out.

- The surfaces obtained may contain numerous peaks or holes (bulls eyes) around the known data. The spatial arrangement of the samples is not taken into consideration: isolated points and clusters have the same importance.

The power used in Equation III.8.18 can assume different values depending from the real-life case (de Mesnard, 2013) and according to that the value used in this work is 2. The IDW method is widely used for the pollutants spatialization (Hart et al., 2009) also for the assessment of human exposure to pollution (Hooyberghs et al., 2006)

Error

All the proposed spatial expansion models were compared by calculating the normalized mean square error (NRMSE) estimating the values of the pollutant concentration, $ce_{i,j}$ in all the N cell were a measured value $c_{i,j}$ was available. Finite periods including T time values j values, it is:

$$NRMSE = \sqrt{\sum_{j=1}^{T}\sum_{i=1}^{N} n_{i,j}\left(\frac{ce_{i,j}-c_{i,j}}{c_{i,j}}\right)^2 / \sum_{j=1}^{T}\sum_{i=1}^{N} n_{i,j}} \quad \text{(III.8.19)}$$

where $n_{i,j}$ assumes value 1 if the measured value $c_{i,j}$ exists, else it is 0.

Results and discussion

One of the result obtained from the data analysis using the MISOE model is the local adjustment coefficient on the area considered. The coefficients were reported directly on the map dIIIided in square cells by means of a chromatic scale for two of the seasons considered (winter and summer in the specific case) for PM10. The data considered do not include values for the weekdays of Saturday and Sunday. In fact, the ROM network does not work on weekends and on other holidays, when the vehicles do not travel (like the 1st of January).

As it can be seen in Figure III.8.2a, in the winter season, the coefficients relating to PM10 provide higher values in the city center and decrease moving towards the periphery of the town. This indicates the presence of a background pollution level that could be represented by domestic heating which, going towards the city center, is also affected by the presence of other activities. The city center, in the specific case of Milan, is affected by heavy daily traffic during working days and therefore favorable to the production and accumulation of pollutants. Nevertheless, the presence of some green areas located in the city center allows to have lower pollution levels and consequently lower adjustment coefficients in some cells. Analyzing the

map of the adjustment coefficients in the summer season (Fig.III.8.2), it can be seen that, unlike in the winter season, there is no prevailing trend in the spatial distribution of the coefficient values. These assume lower values than in winter for the majority of the cells, and few hot spots can be identified. These spots correspond to ground cells, such as the Navigli, characterized by intense summer activities which correspond to high vehicular traffic intensity determining the consequent increase in the local pollution levels. In general, the orography of Milan, similarly to the whole Po plane, is not favorable to the dispersion of pollutants and are characterized by the tendency of airborne pollutants to stagnate in the atmosphere, showing high measured concentrations.

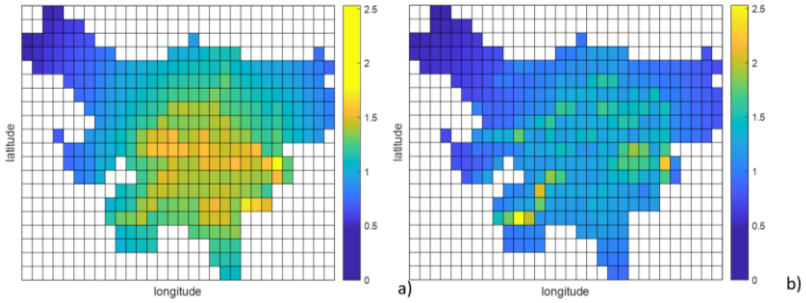


Figure III.8.2. *Map of local monthly adjustment coefficients for PM10 for all the city cells: a) winter season, b) summer season.*

The whole MISOE model developed was applied to the PM10 values measured by the ROM network in Milan considering days with different climatic and vehicular traffic conditions. In particular, four days of January 2020 (Figures III.8.3 and III.8.4) are considered as an example of winter conditions and one day of May 2020 as an example of summer conditions (Figure III.8.5). The maps obtained show the measured values as circles while the estimated values are represented by triangles. PM10 concentrations are reported on a red scale if they are distributed in a high concentration range and on a green scale if they are distributed in a low concentration range. Both days whose corresponding data are shown in Figure III.8.3 were characterized by a high vehicular traffic which considerably contributed to the measured PM10 values. The map shown in Figure III.8.3a, shows a rather homogeneous distribution of concentrations, in fact the low wind speed (32.4 km/h from 202.5°) has favoured the pollutants stagnation and accumulation. In Figure III.8.3b, the distribution of concentrations is more affected by the influence of a slight wind which contributes little to the dispersion of pollutants while maintaining high concentration values.

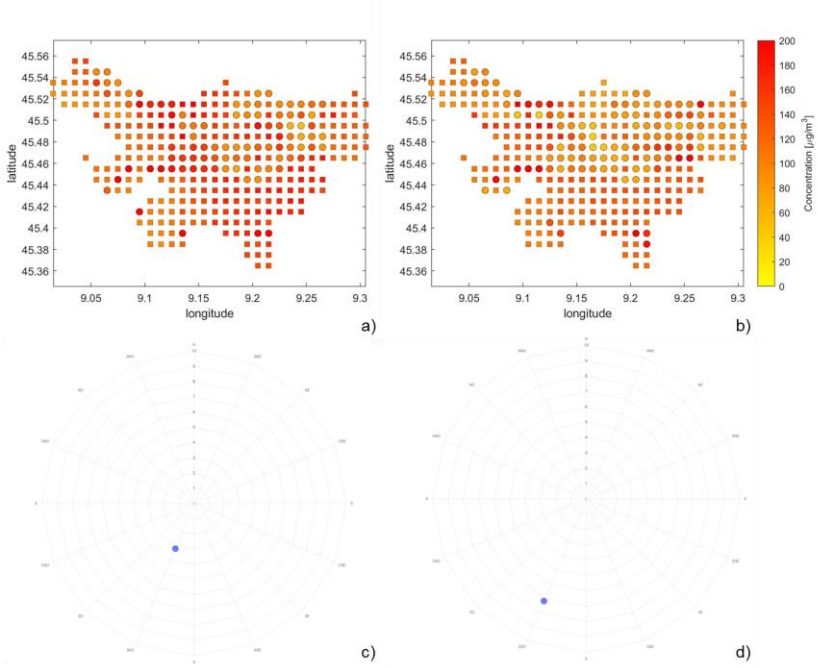


Figure III.8.3. *PM10 daily average concentrations measured in Milan by the ROM network during a) 13rd (wind of 3.2 km/h from 202.5°) and b) 17th (wind of 6.5 km/h from 247.5°) January 2020. Panels c) and d) report the wind intensities and directions.*

Rain affects pollution by breaking it down as can be seen from Figure III.8.4a. The particles of the solid pollutant are captured by the rain drops and brought down to the ground (wash-out phenomena). The PM10 values shown in Figure III.8.4 are low as they represent rainy days (Fig.III.8.4a) and windy days with high intensity (Figure III.8.4b) despite moderate vehicular traffic intensity. The pollution mitigation effect appears to be greater in the presence of rain than in high windy days.

The spring/summer period considered is May, which in this case recorded higher wind intensity values than January, so the values measured are lower than January. Moreover, the stop to the domestic heating contributed to the decreasing of pollutants concentrations. In the representative map of 28/05 shown in Figure III.8.5, it can be seen how the dispersion of PM10 is aligned along the wind direction (SSW). The medium-high wind speed and the high vehicular traffic intensity determine PM10 daily averages very high are in some places.

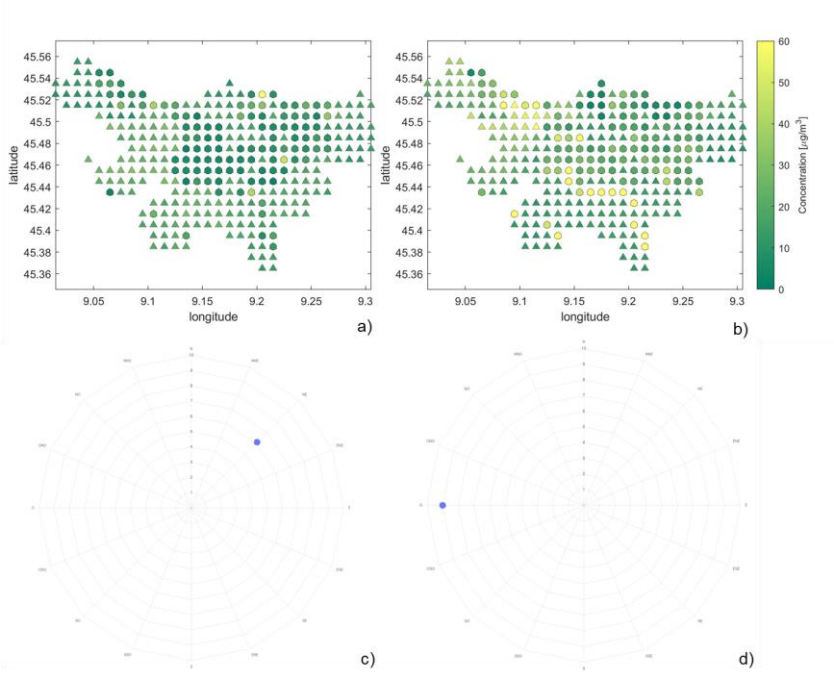


Figure III.8.4. *PM10 daily average concentrations measured in Milan by the ROM network during a) 20th (wind of 6.1 km/h from 45° with rain) and b) 29th (wind of 9 km/h from 270°) January 2020. Panels c) and d) report the wind intensities and directions.*

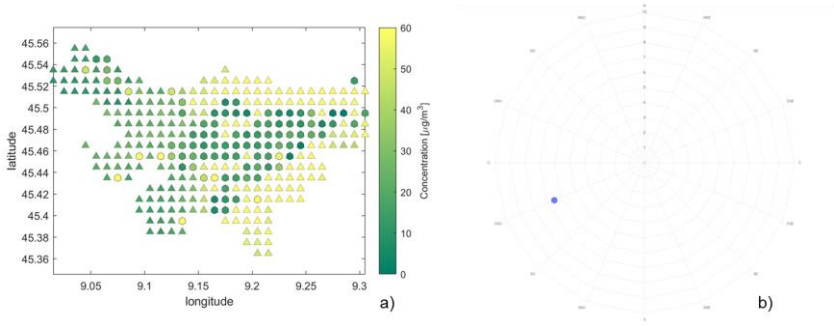


Figure III.8.5. *PM10 daily average concentrations measured in Milan by the ROM network during 28th May 2020 (wind of 7.2 km/h from 202.5°). Panel b) reports the wind intensity and direction.*

The NRMSE results are reported in Table III.8.1. They clearly indicate that the modified isoellipse model (MISOE) has the lowest values. In general, the errors decrease as the wind intensity decreases because in the presence of high wind intensity, the dispersion increase and the estimates are affected by the greater dispersion. At low wind intensity values, the errors

associated with the IDW and the ISOE models are similar and are slightly higher than the error associated with the MISOE model. The ISOE model allows to have a better estimate of the missing data as it contains the information of the historical pollution of the cell.

Table.III.8.1. *NRMSE for the model analyzed.*

		NRMSE		
	N° of measaured data	**ISOE**	**MISOE**	**IDW**
January	2265	0.4	0.17	0.45
May	2223	0.35	0.12	0.4
January wind <1.5 m/s	729	0.25	0.16	0.3
January wind >1.5 m/s	1536	0.33	0.14	0.35
May wind <2.5 m/s	1182	0.42	0.2	0.4
May wind >2.5 m/s	1041	0.57	0.25	0.6

From the maps obtained on a daily basis it can be seen how the local adjustment coefficients vary according to the conditions (Fig. III.8.6). On some days, pollution conditions located in the northern part of the city are noted. In particular, the day represented in Figure III.8.6 is the one that shows the highest values of the local adjustment coefficients. Figure III.8.6 illustrates a localized pollution in a specific cell due to local conditions that increase the pollution levels compared to the rest of the city. Considering the cell relating to the city center as shown in Figure III.8.2, it can be seen that it assumes medium-high values throughout the month. The same behaviour assumes the cell relating to the industrial area. As regards the cell located on the north-western periphery, the local adjustment coefficients assume medium-low values throughout the month of January. The area with high vehicular traffic intensity has, especially during Friday high coefficients which indicate the high actIIIity of the area.

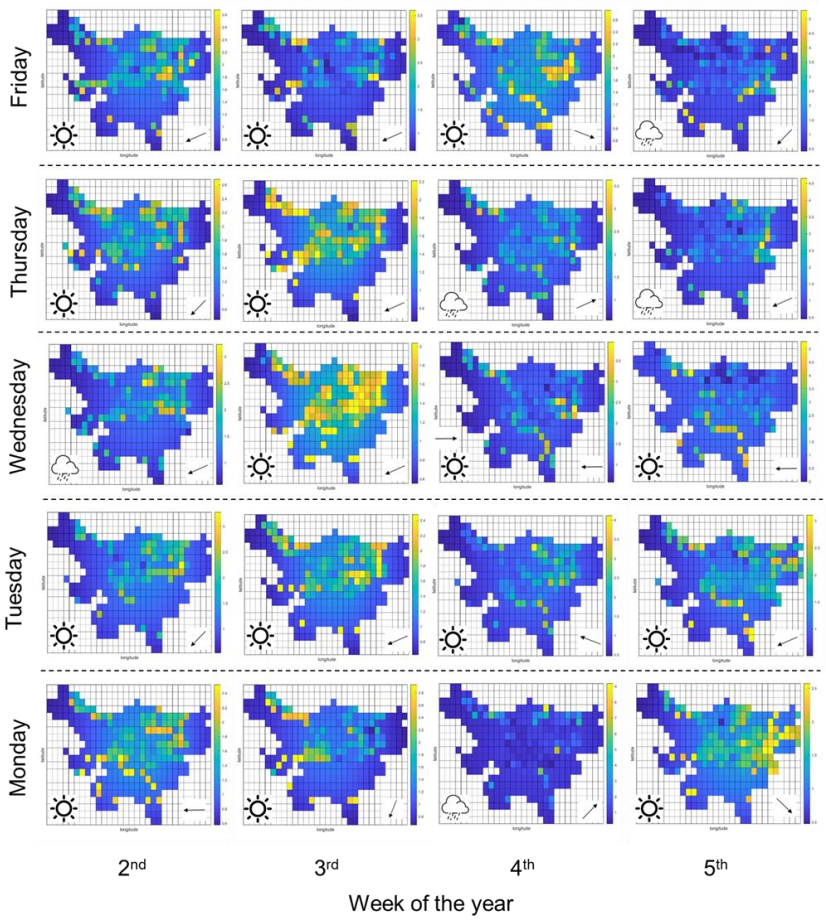


Figure III.8.6. *Map of the local adjustment coefficients calculated on a daily basis for the month of January 2020. Dashed rectangles indicate days with available measurements.*

Given the high spatial resolution of the system considered, it was possible to analyze the pollutants trends, in particular of PM10, for some particularly interesting cells (reported in Fig. III.8.7). Four cells representative of the territory have been identified. The first cell considered, A, represents the city center and considers all those activities mainly related to tourism and vehicular traffic due to public transport. A second interesting cell, B, is the one representing the industrial area and a large number of manufacturing activities. Further cells considered were a representative of the peripheral area located to the NW of the city, C, and a cell characterized by a high flow of vehicular traffic, D.

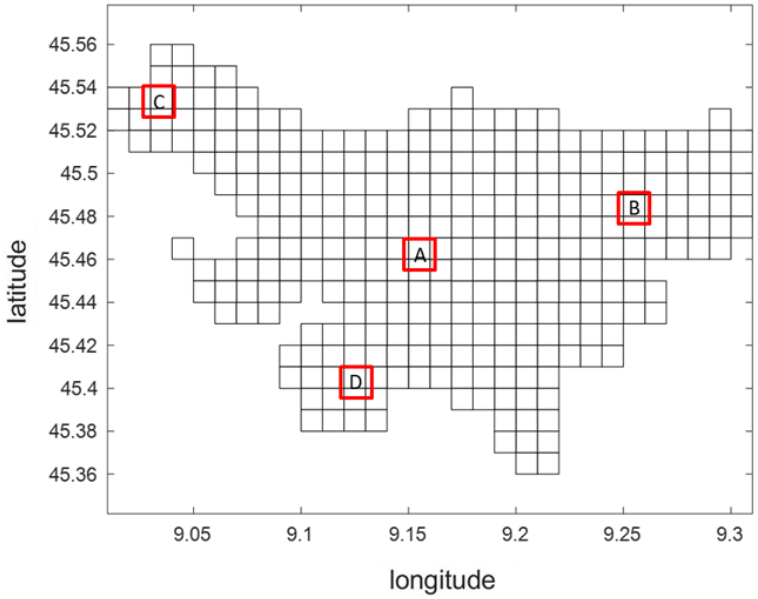


Figure III.8.7. *Detail of the cells used for the analysis.*

Figure III.8.7 shows the trends of the local adjustment coefficients calculated on a monthly basis for the four cells considered in Figure III.8.7. It is noted that in all cases a minimum is observed in the month of March. This result is in line with what derives from the analysis of the pollution trend during the lockdown (Lotrecchiano et al., 2021), in particular between the month of January and March there is an average decrease in the local adjustment coefficients of 70%. The coefficients are on average lower values in May than in January, except in the area high vehicular traffic intensity.

The variation of PM10 daily averages for the different cell considered are reported in Figure III.8.9 the months of January and May 2020. The concentration values relative to January are generally higher than those taken in May, in agreement with seasonal pollution trends. In January the high pollutants level related to the cell C (Fig. III.8.9d) is clear, and these levels remain high even during the month of May. This trend agrees with the high value of the local adjustment coefficient shown in Figure III.8.2 which indicates a characteristic high pollution level in this cell for both months. The cell B, representing the city center (Fig. III.8.9a) shows more or less constant levels in both January and May. As Milan is city attracting tourists, the center is always subject to multiple activities that affect PM10 levels. Furthermore, May is the month in with the greatest touristic presence producing high vehicular traffic which maintain high levels of pollution.

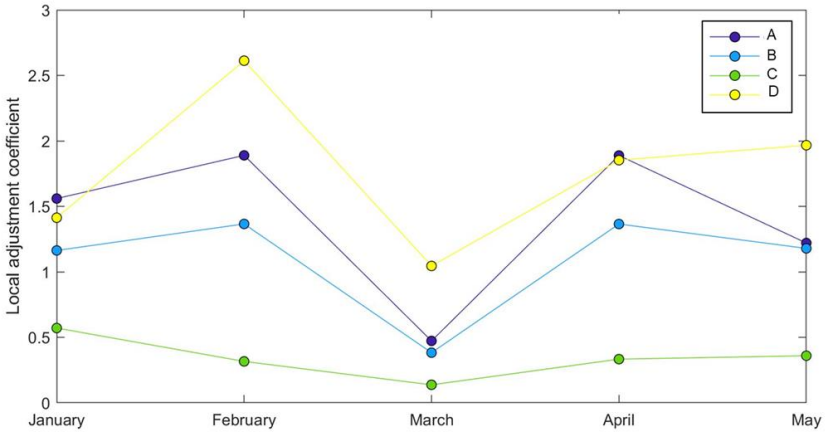


Figure III.8.8. *Local adjustment coefficients for the A-D squares considered in Figure III.8.7 for January-May 2020.*

The cell C, representative of the north-west periphery (Fig. III.8.9c) shows, also in agreement with the local adjustment coefficient, low levels of pollution. The cell B, relating to the industrial zone (Fig. III.8.9b)) has higher levels in winter than in spring. In Figure III.8.9, in the week between 20 and 24 January, the pollutants accumulation due to the persistence of high pressure and the absence of wind or rain is evident. As regards January 20, the values shown in Figure III.8.9 are also evident in Figure III.8.6. On a weekly basis it is observed that during Fridays the cell is always greater than the other cells, this highlights both the flow of cars leaving the city for the weekend and greater circulation due to the afternoon-evening activities that characterize the area.

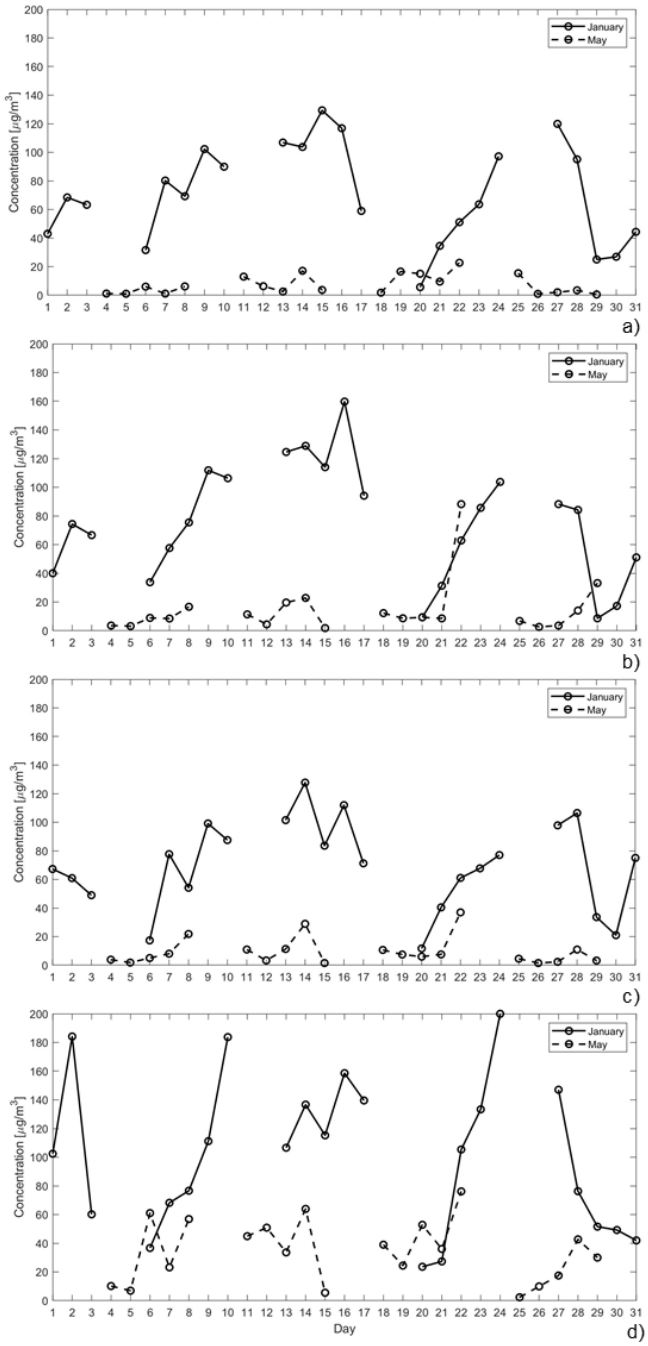


Figure III.8.9. *Daily PM10 average concentrations for A) the city center, B) industrial zone, C) periphery NW, D) high traffic zone considered in Figure III.8.5 for January 2020 and May 2020.*

Conclusions

It was possible to integrate data from a dynamic road network (ROM) in which pollution measurement sensors were mounted over delivery vans driving in Milan. Different integration methods were tested. Purposely developed ISOE and MISOE models were compared with the IDW method known in the literature. In windy days, the ISOE model that that accounts for the effect of wind provides lower error values than the IDW, a traditional method, in which only distances are accounted for. The MISOE model accounting also for local specific deviations of pollution source provides more accurate estimated values with a low associated error. In fact, in this method, the introduction of the local adjusting coefficients is able to take into account the local characteristics of the pollution on the basis of historical trends, have allowed the improvement of the performance of the ISOE model, for which the MISOE model returns estimated values that are much more similar to those effectively measured. This possibility maximizes the potential of pollution measurement networks characterized by high space and time resolution. Furthermore, the analysis of the distribution of the local adjusting coefficients can be useful the most critical areas with respect to pollution level and may be of help to policy makers for the implementation of pollution mitigation policies.

III.9. Application of methods for reducing the dimensionality and classification of a dataset containing information on air quality

Air quality data

The data used concern the average daily concentrations measured by the ARPALombardia air quality network in Milan during 2013-2021. The pollutants considered were only those with the complete data set, namely: PM10, PM2.5, NO_2, SO_2, O_3, NH_3, Black Carbon (BC) and Benzene (C_6H_6). The trends of the pollutants in the period considered are shown in the following Figures (Fig. III.9.1):

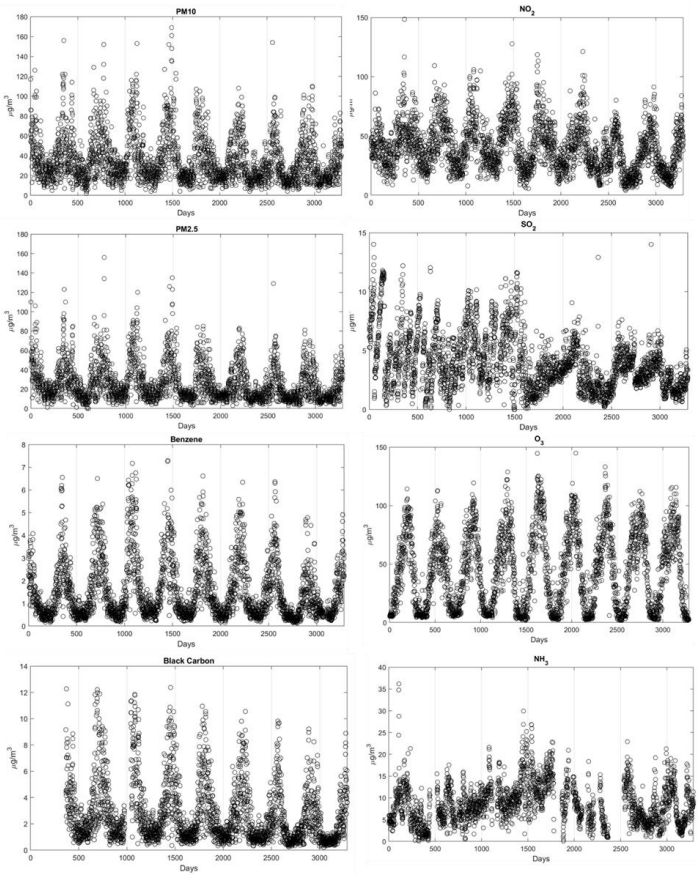


Figure III.9.1. *Pollutant temporal series 2013/2021.*

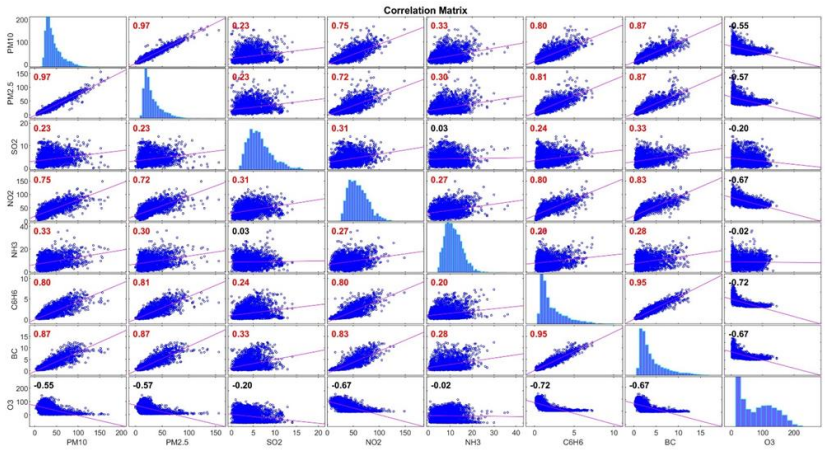


Figure III.9.2. *Pollutants correlation matrix.*

As can be seen from the correlation matrix shown in Figure III.9.2, some pollutants are strongly correlated with each other. PM10, PM2.5 and BC have a correlation in the range 0.8-0.97, being part of the same unit; in fact, PM2.5 and BC are contained within PM10. Another strong relationship is present between BC and benzene (C_6H_6) which show a correlation value of 0.95; being both carbonaceous compounds, they tend to assume the same behavior and therefore to present a strong linear relationship between them. The same happens between benzene and particulate matter PM10 and PM2.5. Correlation coefficients highlighted in red indicate which pairs of variables have significantly non-zero correlations, so for these time series all pairs of variables have significantly non-zero correlations.

Input matrix development

An X input matrix has been constructed in which rows correspond to observations and columns to variables. Vector y is a column vector of class labels for each observation in X. The X matrix is made up of the 8 pollutants considered throughout the selected period of size 3287*8. To define the vector y, the air quality index (AQI) was calculated in order to create an index that could take into account all pollutants. This index is calculated for each day (d) and for each pollutant (i) according to the relation 5.9.1 in which C_m is the average daily concentration measured for the pollutant i considered and C_{lim} is the limit concentration reported in Legislative Decree 155 / 2010.

$$AQI_{d,i} = \frac{C_m(d,i)}{C_{Lim}(i)} \qquad \text{(III.9.1)}$$

The AQI is therefore defined as the sum of the $AQI_{d,i}$ calculated. Therefore, three classes were identified into which the AQI index was divided:

- If AQI <= 2, class = 0 - absence of pollution
- If 2 <AQI≤4, class = 1 - pollution to the limit
- If AQI> 4, class = 2 - high pollution

The pollutants values included in the input matrix have been standardized to make them homogeneous subtracting the average and dividing for the standard deviation. Starting from the original sample, and having known the mean and standard deviation of the populations to which the variables belong, standardization allows the variables normalization and to obtain a new sample, which will have zero mean and unit variance.

Backward feature selection

The Backward Feature Selection is part of the wrapping methods. They are able to evaluate the importance of each feature in models that present a subset of features until the optimal subset is obtained. In wrapper methods, the feature selection process is based on a specific machine learning algorithm that tries to fit on a given dataset. It follows a greedy search approach by evaluating all the possible features combinations against the evaluation criterion. This criterion is simply the performance measure which depends on the problem type, (for e.g. for regression evaluation criterion can be p-values, R-squared, Adjusted R-squared, similarly for classification the evaluation criterion can be accuracy, precision, recall, f1-score, etc). Finally, it selects the combination of features that gives the optimal results for the specified machine learning algorithm. In general, the algorithm begins with a model that contains all variables under consideration (called the Full Model), then starts removing the least significant variables one after the other until a pre-specified stopping rule is reached or until no variable is left in the model

The Backward Feature Selection, starts considering all features and remove the least significant feature at each iteration, improving model performance until no improvement is observed on removing features. This algorithm has been implemented in Matlab using the *sequentialfs* function which refers to the ECOC model (error-correcting output codes). The function selects a subset of features from the X data matrix that best predicts the data in y by sequencing the features until there is no improvement in the prediction. Then the vector y containing the classes into which the AQI was divided, was fed together with the values of the pollutants to the *sequentialfs* function. A K-fold cross validation partition has been implemented through the *cvpartition* function considering:

- NumObservations: 3287

- NumTestSets: 10
- TrainSize: 2959 2958 2958 2958 2958 2958 2958 2958 2959 2959
- TestSize: 328 329 329 329 329 329 329 329 328 328

Following the application of the backward feature selection algorithm, the features inherent to PM10, PM2.5 and benzene are significant. Considering only these pollutants, the loss was calculated and found to be 0.024. A sensitivity analysis was carried out by varying the number of test sets and the algorithm always returns the same features as significant.

Principal Component Analysis (PCA)

Principal component analysis allows to derive the vectors that represent most of the data variance. The reduced set of linear orthogonal projections (λ) is obtained by linearly combining the original variables in an appropriate manner. The PCA function implemented in Matlab returns the principal components of the input matrix. A first criterion for choosing the number of main components is to take a number of components equal to the number of eigenvalues that are able to cover 80-90% of the total variability. Therefore, as indicated in the following table (Tab.III.9.1), which shows the percentages of each λi, it is noted that analysing the percentage variability explained by the main components, the first component explains 63.2% of the variability while the second 13% and the third 11.2 % (Tab. III.9.1). Considering the cumulative percentage (Tab. III.9.1) it is clear that the first three components explain 87.5% of the total variability. Therefore, the principal components to consider are three. The values of the eigenvalues of all the principal components are made explicit by the latent function (Tab. III.9.1).

Table III.9.1. *Components of PCA.*

	Principal components							
	λ1	**λ2**	**λ3**	**λ4**	**λ5**	**λ6**	**λ7**	**λ8**
Variability %	63.2	13	11.15	6.2	3.2	2.5	0.45	0.3
Cumulative %	63.2	76.2	87.35	93.55	96.75	99.25	99.7	100
Eigenvalues	5.05	1.04	0.9	0.5	0.26	0.2	0.03	0.02

Following Kaiser's rule, only those components are considered that have an eigenvalue greater than or equal to one, or, equivalently, the components that have variance greater than the average (obtained as the average of the λi). Therefore, considering the eigenvalues shown in table III.9.1, the components to be considered are two. The choice of the number of components sufficient to reproduce the starting data with a good

approximation can also be made through the analysis of the eigenvalue graph or scree plot. The scree plot shown in Figure III.9.3 shows the eigenvalues of the main components and it can be seen that the components to be considered are the number of components corresponding to the elbow point of the broken line, therefore three.

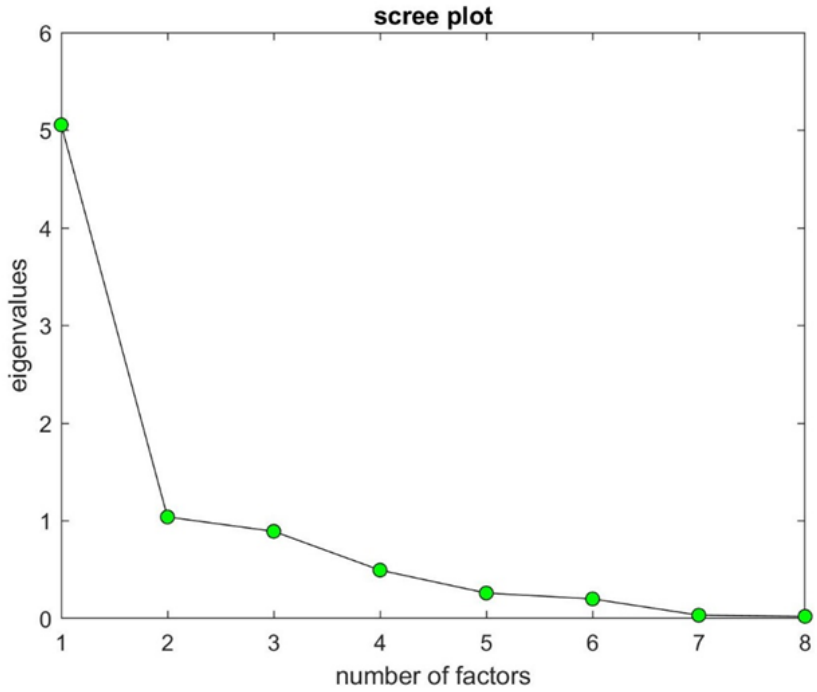


Figure III.9.3. *Scree plot.*

The PCA is completed by analyzing how the individual variables affect the component and then the coefficients returned by the function are analyzed. These coefficients are shown in table III.9.2.

The biplot in Figure III.9.4 shows all eight components in the form of a vector whose direction and length indicate how much each variable contributes to the two main components. The analysis of the variability shows that the first three main components are able to explain 87.4% of the variability of the dataset. From the analysis of the plot, it can be seen that the variable that contributes most to the main components is v5, followed by v1, v2, v6 and v7. The first principal component has positive coefficients for all variables except v8 whose coefficient is negative. Variables v1, v2, v6, and v7 have the largest coefficients for the first principal component. The second main component has the higher coefficient for v5 and negative coefficients for v3, v4 4 v6. The v7 has zero coefficient along the second main

component. Variable 3 contributes predominantly to the third main component which has low negative coefficients for v1, v2, v6, v7.

Table III.9.2. *Principal components coefficients.*

		Principal components							
Variables		**λ1**	**λ2**	**λ3**	**λ4**	**λ5**	**λ6**	**λ7**	**λ8**
v1	**PM10**	0.41	0.15	-0.03	0.43	-0.23	-0.25	0.06	0.71
v2	**PM2.5**	0.41	0.11	-0.05	0.44	-0.33	-0.18	-0.20	-0.67
v3	**SO_2**	0.15	-0.41	0.89	0.01	-0.10	0.05	-0.03	0.02
v4	**NO_2**	0.39	-0.05	0.02	-0.26	0.64	-0.59	-0.12	-0.06
v5	**NH_3**	0.14	0.83	0.34	-0.39	-0.12	0.08	-0.03	-0.01
v6	**C_6H_6**	0.41	-0.08	-0.14	-0.03	0.22	0.64	-0.58	0.13
v7	**BC**	0.43	0.00	-0.04	0.05	0.23	0.36	0.78	-0.14
v8	**O_3**	-0.33	0.33	0.25	0.63	0.55	0.10	-0.06	-0.03

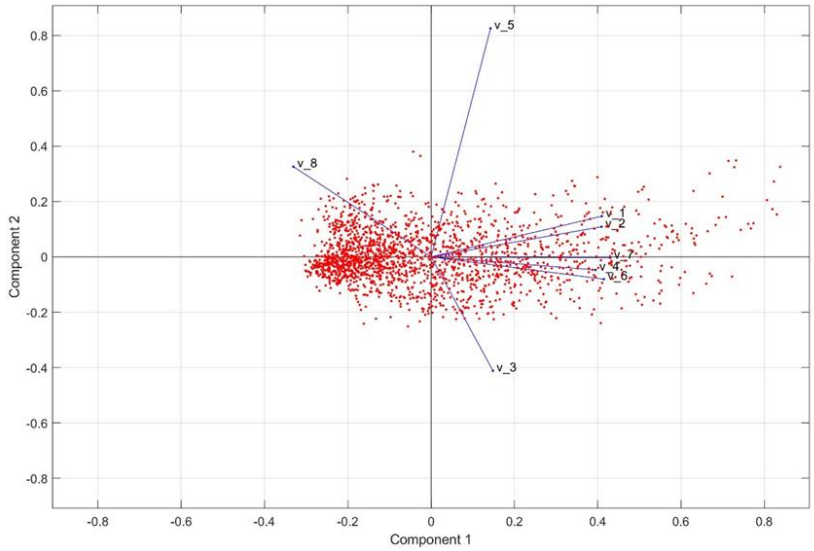


Figure III.9.4. *Variables biplot along the principal components.*

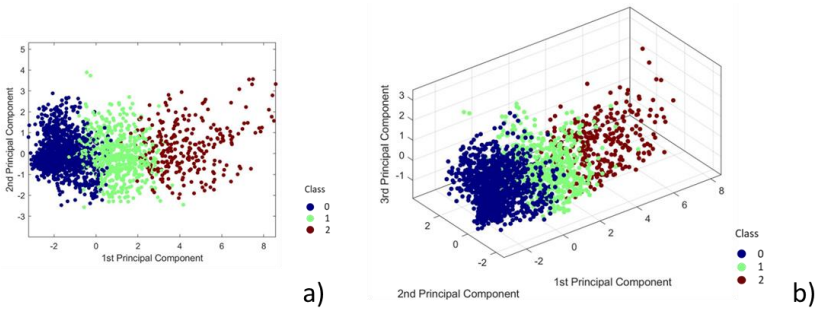


Figure III.9.5. *Data variability along the first a) two and b) three components.*

Placing the data along the three main components, it can be seen that they are more distributed along the first main component for which the greatest variability of the data occurs (Fig. III.9.5). Furthermore, considering the classification of the data, it can be seen that they occur in three different clusters with some overlapping points considering both two main components and three.
In summary, the main components are:

- λ1: mainly takes into account PM10, PM2.5, Benzene and BC. These pollutants are mainly produced by combustion so the first main component takes into account the contribution of traffic and domestic heating.
- λ2: mainly takes into account NH_3. Ammonia reacting in the atmosphere, produces by-products that are present in the particulate matter. Therefore, the second main component still takes into account road transport but also the agricultural component which is the major source of ammonia.
- λ3: mainly takes into account SO_2. Its presence in the atmosphere is mainly due to the combustion of fossil fuels and the main anthropogenic sources are systems for heating and energy production fueled by diesel, coal and fuel oils. SO_2 is typically produced by aircraft and ships whose fuels are sulfur-based. In this case, it indicates the influence of the nearby Linate (MI) airport.

t-distributed stochastic neighbour embedding (t-SNE)

t-distributed stochastic neighbour embedding (t-SNE) is a dimensionality reduction algorithm developed by Geoffrey Hinton and Laurens van der Maaten, widely used as a machine learning tool in many research fields (van der Maaten and Hinton, 2008). It is a non-linear dimensionality reduction technique that lends itself particularly to the embedding of high-dimensional datasets into a two- or three-dimensional space, where they can be visualized

via a scatter plot. The algorithm models the points so that nearby objects in the original space are close in the reduced-dimensional space, and distant objects are far away, trying to preserve the local structure. In the first part of the algorithm, a probability distribution function is constructed which associates a high probability value if the two points are similar, low if they are dissimilar to each pair of points in the original high-dimensional space. Then a second analogous probability distribution function is defined, in the reduced-dimensional space. The algorithm therefore minimizes the Kullback-Leibler divergence of the two distributions by descending the gradient, reorganizing the points in the reduced-size space. Given a set of N objects $x_1, \cdots . x_N$ in a space with a high dimensionality, the t-sne algorithm define a probability distribution p_{ij}, symmetric in the two variables and proportional to the similarity between x_i and x_j, defined as follows:

$$p_{ij} = \frac{p_{j|i} + p_{i|j}}{2N} \tag{III.9.2}$$

with $p_{ij} = p_{ji}$, $p_{ii} = 0$, and $\sum_{i,j} p_{ij} = 1$
where

$$p_{j|i} = \frac{\exp(-\|x_i - x_j\|^2 / 2\sigma_i^2)}{\sum_{k \neq i} \exp(-\|x_i - x_k\|^2 / 2\sigma_i^2)} \tag{III.9.3}$$

and set $p_{i|i} = 0$ and $\sum_j p_{j|i} = 1$ for all *i*

The Gaussian bandwidth σ_i is chosen in such a way that the distribution perplexity is equal to the perplexity given as the algorithm hyperparameter. This condition allows the width adjusts to points density, with lower σ_i values in high density areas. t-SNE tries to reconstruct a d-dimensional map $y_1, \cdots . y_N$ (with $y_i \in \mathbb{R}^d$)where points represents the similarity p_{ij} in the original space. So, the similarity q_{ij} between two points y_i and y_j in the space with reduced dimensionality is defined by Equation III.9.4. In detail, for $i \neq j$, define q_{ij} as :

$$q_{ij} = \frac{(1 + \|y_i - y_j\|^2)^{-1}}{\sum_{k \neq m} (1 + \|y_k - y_m\|^2)^{-1}} \tag{III.9.4}$$

and set $q_{ii} = 0$. The position y_i in the space with reduced dimensionality is calculated by minimizing, using the gradient descend, the non-symmetric Kullback-Leibler divergence of the Q distribution respect the P distribution that is:

$$KL(P||Q) = \sum_{i \neq j} p_{ij} \, log \frac{p_{ij}}{q_{ij}} \quad \text{(III.9.5)}$$

Using the Kullback-Leibler divergency as the objective function allows to have higher hinders is nearby points in the original space (high p_{ij}) are considered far in the reduced dimensionality space (q_{ij} small) while the opposite has less influence, preserving the points distribution local structure. In this way it is possible to obtain a map that reflects the similarities between the high-dimensional inputs (van der Maaten and Hinton, 2008).

The *tsne* function comes with a list of hyperparameters that can be suitably modified. Once the space for the t-SNE was defined, a sensitivity analysis was carried out to optimize the model parameters and increase model performance. The optimized parameters were perplexity, learning rate and exaggeration. The parameter for the dimension of the lower space *NumDimension* to which convert the dataset was chosen between two and three because it is necessary to display the dataset in a two-dimensional or three-dimensional space. The loss calculated by the Matlab function represents the Kullback-Leibler divergence which is 1.07 in 2D embedding and 0.97 in 3D embedding as can also be seen from the Figures below (Fig.III.9.6):

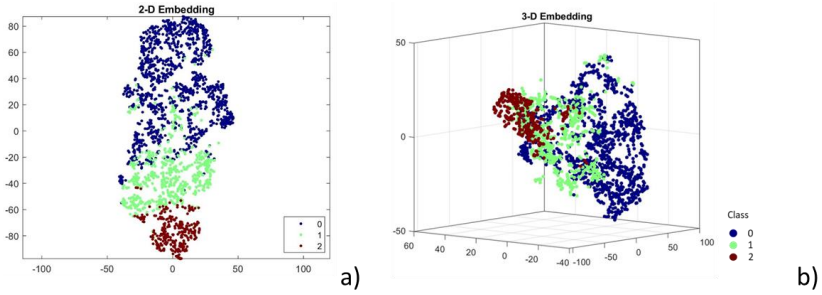


Figure III.9.6. *a) 2D and b) 3D embedding visualization.*

Although the loss is lower in 3D space, for simplicity's sake it is chosen to work in 2D space considering that the loss value obtained is not very different from the relative 2D. Subsequently, different methods were compared to calculate the divergence, in addition to the already considered Kullback-Leibler divergence.

Perplexity

The first parameter to be changed is perplexity. A perplexity represents a target number of neighbors for the center point. In general, the greater the perplexity, the greater the variance value. As perplexity increases we will get

more stable configurations, however the value of this parameter should always be less than the total number of points in order to avoid the algorithm failing and having unexpected behavior. This parameter is set by default equal to 30 and the t-SNE has been calculated by setting it equal to 100, 50, 5 (Fig. III.9.7). Although the recommended perplexity values are in the 5-50 range, it was decided to investigate what happens even at very high perplexity values.

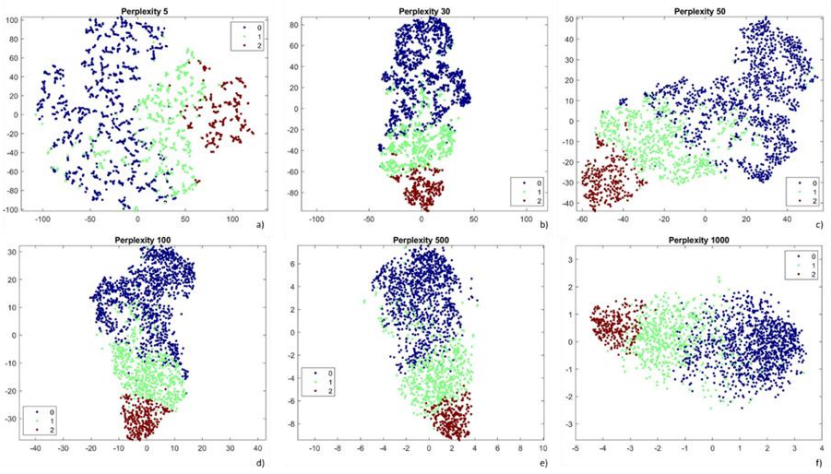


Figure III.9.7. *t-SNE obtained considering different perplexities.*

Increasing the perplexity, the clusters are less dense and have few points of overlap. Furthermore, by analyzing the loss value, there is a significant decrease up to the value of 0.12 for the perplexity equal to 1000. Although the loss value decreases considerably as the perplexity increases, the clusters are less separable. Therefore, considering the loss value and the separation of the clusters, it was concluded that a perplexity value equal to 30 is the right compromise between the two aspects (loss = 0.97).

Learning rate

If the learning rate is too small, t-SNE could converge to a local minimum. If, on the other hand, it is very large, the optimization may initially have a value of the Kullback-Leibler divergence that increases instead of decreasing. For this reason, it is important to set the correct learning rate value in order to have a good model performance. The default learning rate value is 500 and usually the range varies between 100 and 1000. Analyzing the results as the learning rate varies, it can be seen that in correspondence with some learning rate values, the clusters are further divided into groups (Fig. III.9.8). This occurs to a lesser extent with a learning rate of 500.

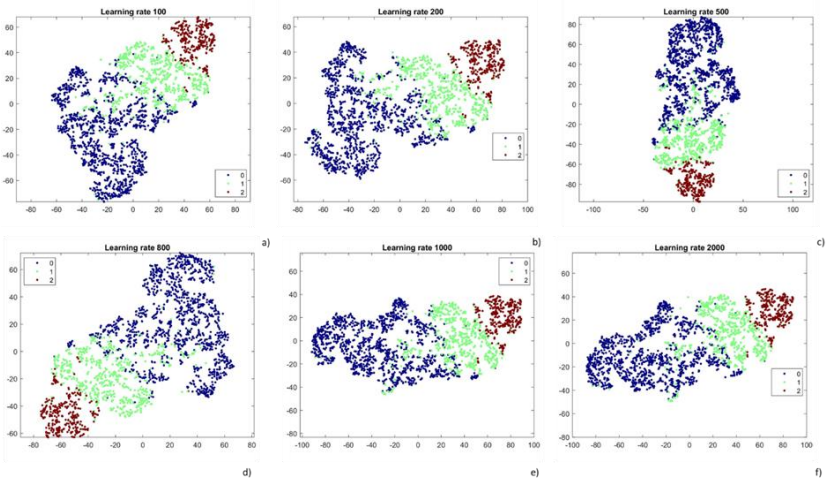


Figure III.9.8. *t-SNE obtained considering different learning rates.*

Exaggeration

The Exaggeration parameter controls how tight the natural clusters are in the original space and reduced space and how much space there will be between them. For larger values, the space between the natural clusters will be greater in the reduced space. Again, the choice of this parameter is not very critical. If the cost function increases during pre-tuning, the initial exaggeration factor or learning rate may be too high.

A large value of exaggeration causes t-SNE to create relatively more space between clusters in Y. If the value of the Kullback-Leibler divergence increases in the initial phase of optimization, it is necessary to try to reduce the exaggeration. In this case the exaggeration value has been set to 4 (Fig. III.9.9).

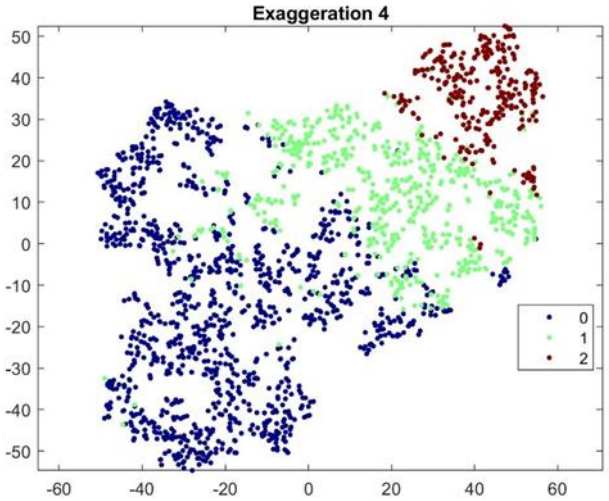


Figure III.9.9. *t-SNE obtained with Exaggeration=4.*

The generally used distance is the Kullback-Leibler divergence as previously mentioned. But it is possible to consider other functions to evaluate the divergence such as the Mahalobis, Cosine, Chebychev and Euclidean distances (Fig. III.9.10)

The Cosine, Chebychev and Euclidean distances show a better classification even if they have some overlapping clusters. The Euclidean distance seems to have a better classification, with the identification of the three clusters and a few overlapping points. However, compared to the Kullback-Leibler distance there is not a noticeable improvement so we choose to keep this distance. As for the Mahalanobis distance, it turns out to be the worst of the four considered. The results evident from the graphs (Fig. III.9.10) are confirmed by the calculated loss values.

So ultimately, the implemented t-SNE has the following characteristics:

- Num dimensions: 2;
- Perplexities: 30;
- Learning rate: 500;
- Exaggeration: 4;
- Kullback-Leibler divergence

For the classification, for homogeneity with the datasets obtained following the Backward feature selection and the PCA, the dataset was obtained considering 3 dimensions and leaving the other characteristics unchanged (perplexity 30, learning rate 500, exaggeration 4).

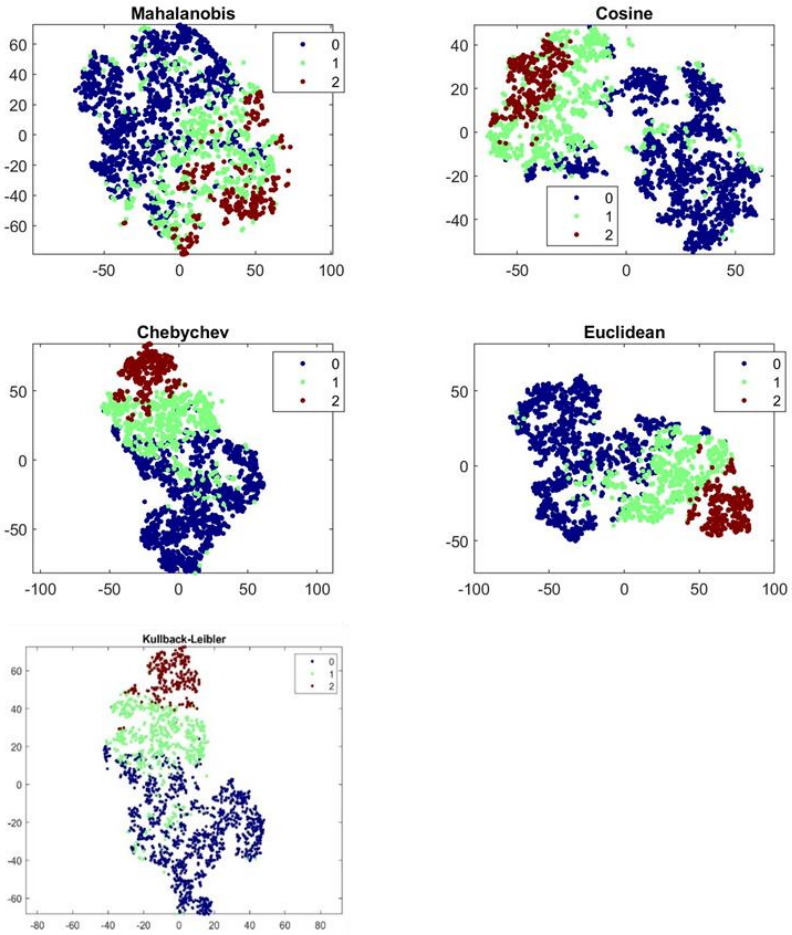


Figure III.9.10. *t-SNE obtained considering different divergences (Mahalanobis, Cosine, Chebychev, Euclidean) and compared with the Kullback-Leibler divergence.*

Definition of a neural network for the data classification

Following the reduction in dimensionality brought about by the previously analyzed algorithms, the reduced dataset is subject to classification according to appropriate algorithms. The following input datasets were used:

- D1 - obtained through Backward feature selection. Pollutants considered: 3 (PM10, PM2.5, C_6H_6)
- D2 - obtained by PCA. Main components used: 3
- D3 - output of the t-SNE procedure. Components used: 3.

The classification was implemented through the definition of three-class neural networks (class 1, 2, 3). The neural network implemented (Fig. III.9.11) consists of:

- 3 input nodes
- 1/2/5/10/15 hidden layers with 10 neurons
- 3 output nodes
- Neuron activation function: softmax
- Error function: cross entropy (CE)
- Error minimization function: Scaled conjugate gradient backpropagation
- Training set: 70% (2301 values)
- Test set: 15% (493 values)
- Validation set: 15% (493 values)

The size of the training, test and validation set, were chosen by analysing the outputs obtained. In particular, by decreasing the training set size, the classification was worse so, the values previously reported, suggested by the Matlab implemented function, was maintained. Following the definition of the sets size, a sensitivity analysis changing the number of hidden layers was carried out. Analyzing the cross entropy, as the number of hidden layers increase in the test phase, with the dataset D1, it can be seen that is higher with 10 hidden layers (Tab. III.9.3) while the percentage error, evaluated as the difference between the output values and its corresponding expected value in the testing phase it is lower in correspondence with 2 hidden layers. Using dataset D2 as input (Tab. III.9.4), the minor error occurs in correspondence of 5 hidden layers even if the cross entropy value is the highest. Dataset D3 (Tab. III.9.5) has higher percentage error values than the previous datasets, but lower cross entropy values. Using dataset D3 as input, the lowest error values are obtained in correspondence of 15 hidden layers.

Table III.9.3. *Cross entropy and percentage error of the neural network with D1 as input.*

BFS	**Samples**	**Cross Entropy**					**Errore%**				
n° hidden layers		**1**	**2**	**5**	**10**	**15**	**1**	**2**	**5**	**10**	**15**
Training	2301	1.7	1.9	1.4	2.1	1.5	7.3	4.3	4.5	4.3	5.4
Validation	493	4.8	5.3	4.0	5.8	4.3	6.7	5.3	6.5	3.9	5.9
Testing	493	4.8	5.3	4.0	5.8	4.3	7.5	4.7	5.9	6.7	4.9

Table III.9.4. *Cross entropy and percentage error of the neural network with D2 as input.*

BFS	Samples	Cross Entropy					Error %				
n° hidden layers		**1**	**2**	**5**	**10**	**15**	**1**	**2**	**5**	**10**	**15**
Training	2301	1.8	1.6	2.2	2.0	1.6	7.2	7.4	7.9	7.4	8.4
Validation	493	5.0	4.4	6.2	5.6	4.4	7.5	6.7	8.1	7.1	6.9
Testing	493	5.0	4.4	6.2	5.6	4.4	8.0	7.3	5.7	7.1	9.1

Table III.9.5. *Cross entropy and percentage error of the neural network with D3 as input.*

BFS	Samples	Cross Entropy					Error %				
n° hidden layers		**1**	**2**	**5**	**10**	**15**	**1**	**2**	**5**	**10**	**15**
Training	2301	1.2	1.0	1.3	1.2	1.4	13.3	11.3	14.3	13.8	13.1
Validation	493	3.1	2.7	3.5	3.2	3.7	14	12.6	12.2	11.3	11.7
Testing	493	3.1	2.7	3.5	3.2	3.7	11.4	11.6	15.8	12.8	11.0

Using dataset D1, it can be seen that more than 95% of the values are correctly classified in the confusion matrix (between 92.5% and 95.3% - Fig. III.9.11). The best classification is obtained with 2 hidden layers in which class 3 is incorrectly classified only for 1.5% (Fig. III.9.11b). Class 3 is the best in the classification in fact it is classified 100% in 3 of the network configurations implemented.

Figure III.9.11. *Test set confusion matrix for a)5, b)10, c)15 hidden layers using dataset D1.*

In the confusion matrix obtained by analyzing dataset D2, between 90.9% and 94.3% of the values are correctly classified (Fig. III.9.12). In this case, the best classification is obtained with 5 hidden layers in which class 3 is incorrectly classified only for 1.8% (Fig. III.9.12b). Class 1 is classified in the best way: using 1 or 2 hidden layers, (Fig. III.9.12 a, b).

Figure III.9.12. *Test set confusion matrix for a)5, b)10, c)15 hidden layers using dataset D2.*

Following the application of the t-SNE algorithm, the D3 dataset was classified by obtaining the following confusion matrices (Fig. III.9.13). They are correctly classified between 84.2% and 89% of the values (Fig. III.9.13). The best classification is obtained in this case with 15 hidden layers in which class 1 is not correctly classified for 4.9% (Fig. III.9.13e). In this case, class 2 is the worst in classification in all three network configurations implemented, reaching even 26.8% of non-classification with 5 hidden layers.

Figure III.9.14. *Test set confusion matrix for a)5, b)10, c)15 hidden layers using dataset D3.*

To improve the neural network performance and thus increase the number of correctly classified values, it is possible to increase the number of neurons, use a larger training set, or for example use more input variables. In this case, the number of neurons was increased to 20 without significant improvements. In the case of 3 input variables considered, the training set is already at its maximum value since there is no further data available.

The best performance of the neural network, intended as a reduction of the cross entropy, trained using the dataset D1 is obtained after 61 epochs (Fig. III.9.15d) with 10 hidden layers. The best performances of the network for a small number of epochs are obtained for 15 hidden layers (Fig. III.9.15c). As for the D2 dataset, it provides the best performance at 34 epochs with 2 hidden layers (Fig. III.9.16). Using dataset D3 as input to the neural network, the best performance is obtained after 15 epochs with 10 hidden layers (Fig. III.9.17d).

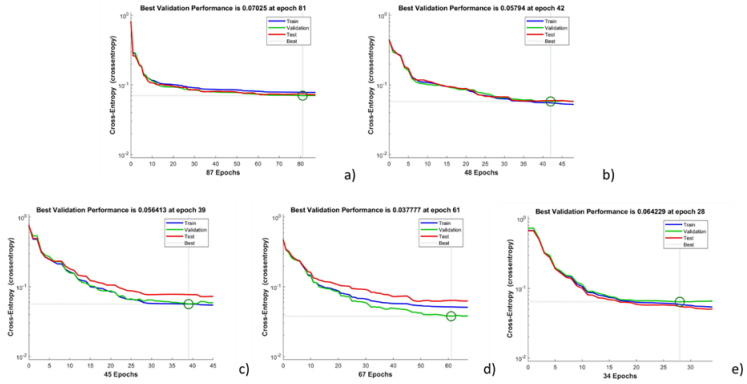


Figure III.9.15. *Cross entropy during the training, validation and test phase of the neural network with input the dataset D1.*

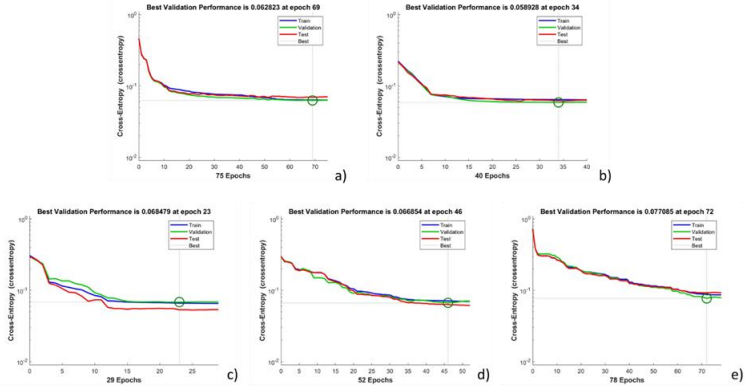


Figure III.9.16. *Cross entropy during the training, validation and test phase of the neural network with input the dataset D2.*

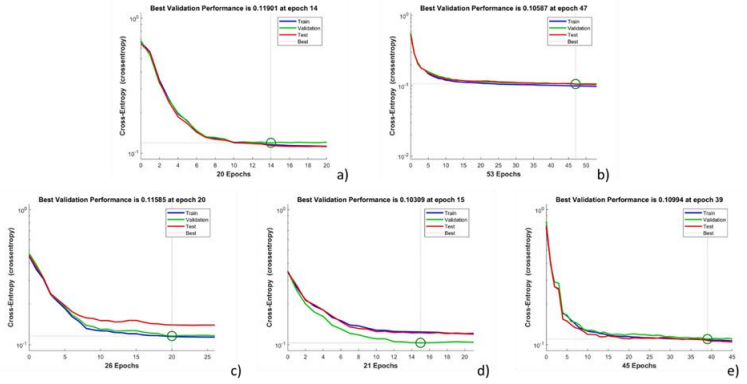


Figure III.9.17. *Cross entropy during the training, validation and test phase of the neural network with input the dataset D3.*

Conclusions

The methods for the reducing dimensionality analysed – Backward feature selection, PCA, t-SNE - reduces the original dataset that includes 8 features to a new dataset with 3 components. Subsequently, the reduced datasets have been used to build an artificial neural network for the classification. Making a comparison between the network performance and the goodness of the classification, it can be stated that:

- dataset D1 is better classified with a neural network with 2 hidden layers with 95.3% of the values correctly classified;
- the D2 dataset is better classified with a neural network with 5 hidden layers with 94.3% of the values correctly classified;
- the D3 dataset is better classified with a neural network with 15 hidden layers with 89% of the values correctly classified.

Among the datasets obtained following the reduction in dimensionality, the best for classification is D1.

Chapter IV
Conclusions and future work

Data obtained experimentally from measurement sensors are limited due to the technological and installation limitations of such sensors. The study conducted in during this work has been aimed at using data beyond the limits of measurement systems, in order to extend them in space and time. The integration of experimental data with a modeling approach was found to be of fundamental importance for the understanding and analysis of the dynamics of pollutants dispersion on a space-time scale.

In a first phase, the characteristics of the monitoring networks were analyzed, defining the correct installation positions of the measurement sensors.

Subsequently, the analysis of the influence of natural phenomena, such as Saharan dust on urban pollution was evaluated.

From the application of the Gaussian dispersion models on an accidental event such as a fire, the dependence of the dispersion of pollutants on meteorological parameters and in particular on the direction and intensity of the wind was highlighted. These considerations were the basis on which to build the model of data spatial expansion.

In fact, the data coming from the real-time on-road monitoring networks (ROMS), although they provide a large quantity of data with a high spatial resolution, leave some points uncovered from the monitoring for which the need arose to develop a model that could estimate the missing data starting from the measured one.

The development of the ISOE wind-dependent spatial expansion model has been improved by adding to it the dependence on historical data measured for each area considered.

These models developed were compared with a model widely used in the literature, such as the inverse distance weighted (IDW), reporting in the comparison better NRMSE values thus showing a good performance of the developed model.

The spatial expansion model could be further improved by including within it the plume Gaussian dispersion which could certainly increase the performance of the model. Finally, a neural network was built for the classification of pollutants following the application of methods for the reduction of dimensionality.

There are many case studies that could be addressed using air quality data and each work carried out highlighted key concepts on the airborne pollutants behavior.

The use of air quality data with a high spatial and temporal resolution can provide detailed information on the environmental matrix that is being considered and allows the implementation of models that well define the trends of pollutants in the air. Furthermore, the analysis of the detailed local pollution levels can be useful for the definition of any areas sensitive to pollution which may be the object of the attention of policy makers for the implementation of pollution mitigation policies.

Chapter III

References

ACI. ACI vehicle park data for Battipaglia municipality, available at: www.comuniitaliani.it

Ahmad, N.A., and Bouhamra, W. (1993) Development of a new long term multiple-source plume model: application on some industrial and residential areas in Kuwait. *Env. Progr.*, **12**, 123-127, https://doi.org/10.1002/ep.670120209.

Al-Ali AR. (2010) A Mobile GPRS-Sensors Array for Air Pollution Monitoring. *IEEE Sens. J.*, **10**, 1666-1671, doi: 10.1109/JSEN.2010.2045890.

AQD, ''Directive 2008/50/EC of the European Parliament'' (2008)

Arnold, M., Seghaier, A., Martin, D., Buat-Ménard, P., Chesselet, R. Géochimie de l'aérosol marin de laMéditerranée Occidentale. In Proceedings of theMediterranean Science *Commission-CIESM. VI Journees Etud. Pollution*, Cannes, France, 2–4 December 1982.

Arroyo, P., Lozano, J., Suárez, J.I., Herrero, J.L., Carmona, P. (2016) Wireless Sensor Network for Air Quality Monitoring and Control. *Chem. Eng. Trans.*, **54**, 217–222, DOI: 10.3390/electronics7120342.

Artaxo P., Martins J.V., Yamaso M.A., Procopio A.S., Pauliquevis T.M., Andrea M.O., Guyon P., Gatti L.V., Cordova Leal A.M. (2002) Physical and chemical properties of aerosols in the wet and dry seasons in Rondônia, Amazonia. *J. Geophysical Res*, **107**, doi:10.1029/2001JD000666.

Artzrouni, M. (2005) Mathematical Demography. In *Encyclopedia of Social Measurement* (Kimberly Kempf-Leonard Eds.), Elsevier, Pages 641-651, https://doi.org/10.1016/B0-12-369398-5/00360-1

Bartier, P.M. and Keller, C.P. (1996) Multivariate Interpolation to Incorporate Thematic Surface Data Using Inverse Distance Weighting (IDW). *Comput. & Geosci.*, **22**, 795-799, http://dx.doi.org/10.1016/0098-3004(96)00021-0.

Beauchamp, M., Malherbe, L., de Fouquet, C., Létinois, L., Tognet, F., (2018) A polynomial approximation of the traffic contributions for kriging-based interpolation of urban air quality model. *Environ. Model. Softw*. **105**, 132–152, https://doi.org/10.1016/j.envsoft.2018.03.033.

Bishop, C. (2006) *Pattern recognition and machine learning*, Springer, Berlin.

Bonasoni, P. Cristofanelli, F. Calzolari, U. Bonafè, F. Evangelisti, A. Stohl, S. Zauli Sajani, R. van Dingenen, T. Colombo, and Y. Balkanski. (2004) Aerosol-ozone correlations during dust transport episodes. *Atmos. Chem. Phys*., **4**, 1201–1215, https://doi.org/10.5194/acp-4-1201-2004.

Brunekreef, B., Holgate, S.T. (2002) Air pollution and health. *Lancet*, **360**, 1233–1242, https://doi.org/10.1016/S0140-6736(02)11274-8.

Bukowiecki, N., Dommen, J., Prèvôt, A.S.H., Richter, R., Weingartner, E., Baltensperger, U. (2003) Fine and ultrafine particles in the Zurich (Switzerland) area measured with a mobile laboratory: an assessment of the seasonal and regional variation throughout a year. *Atmos. Chem. Phys*., **3**, 1477–1494, available at: www.atmos-chem-phys.org/acp/3/1477/.

Bukowiecki, N., Dommen, J., Prèvôt, A.S.H., Richter, R., Weingartner, E., Baltensperger, U. (2002a) A mobile pollutant measurement laboratory—measuring gas phase and aerosol ambient concentrations with high spatial and temporal resolution. *Atmos. Environ.*, **36**, 5569–5579, 10.1016/S1352-2310(02)00694-5.

Bukowiecki, N., Kittelson, D.B., Watts, W.F., Burtscher, H., Weingartner, E., Baltensperger, U. (2002b) Real-time characterization of ultrafine and accumulation mode particles in ambient combustion aerosols. *J. Aerosol Sci.*, **33**, 1139–1154. doi: 10.1016/S0021-8502(02)00063-0.

Buonanno, G., Pacitto, A., Stabile, L., Viana, M., Scungio, M., Reche, C., Querol, X., Alastuey, A., Rivas, B., Álvarez-Pedrerol, M., Sunyer, J., van Drooge, B.L., Grimalt, J.O., Sozzi, R., Vigoa, P. (2018) Particle-related exposure, dose and lung cancer risk of primary school children in two European countries. *Sci. Total Environ.*, **616–617**, 720–729, https://doi.org/10.1016/j.scitotenv.2017.10.256.

Calder, K.L. (1967) Multiple-source plume models of urban air pollution—their general structure. *Atmos. Environ.*, **11**, 403-414, https://doi.org/10.1016/0004-6981(77)90001-4

Canepa, E. (2004) An overview about the study of downwash effects on dispersion of airborne pollutants. *Environ Model. Softw.*, **19,** 1077-1087, DOI: 10.1016/j.envsoft.2003.11.011

Cariou S., Fanlo J.L., Stitou Y., Buty D., Saman, D., Akiki R. (2016) Application of ODEMS (odorant dispersion and emissions monitoring system) to measure odorous emissions from composting plant. *Chem. Eng. Trans.*, **54**, 241–246. DOI: 10.3303/CET1654041.

Castell N., Schneider P., Grossberndt S., Fredriksen M. F., Sousa-Santos G., Vogt M., Bartonova, A. (2018) Localized real-time information on outdoor air quality at kindergartens in Oslo, Norway using low-cost sensor nodes. *Environ. Res.*, **165**, 410-419, https://doi.org/10.1016/j.envres.2017.10.019.

Chiari, M., Lucarelli, A., Nava, S., Paperetti, L., D'Alessandro, A., Mazzei, F., Prati, P., Zucchiatti, A., Marcazzan, G., Valli, G. Aerosol transport

episodes in italian urban environments detected by PIXE analysis. Proc. of the *10th International Conference on Particle Induced X-ray Emission and Its Analytical Applications PIXE*, 4–8 June, 2004, Portoroz, Slovenia,

Collins F.S., Grayand, G.M., Bucher, J.R. (2008) Transforming Environmental Health Protection. *Science*, **319**, 906-907, DOI: 10.1126/science.1154619.

Connolly, P. Gaussian Plume Model in MATLAB. University of Manchester.

Daly, A., Zannetti, P., Echekki, T. (2012) A combination of fire and dispersion modeling techniques for simulating a warehouse fire. International. *J. Saf. & Secur. Eng*., **2,** 368–380, doi: 10.2495/SAFE-V2-N4-368-380.

de Hartog, J.J., Hoek, G., Peters, A., Timonen, K.L., IbaldMulli, A., Brunekreef, B., Heinrich, J., Tiittanen, P., van Wijnen, J.H., Kreyling, W., Kulmala, M., Pekkanen, J. (2003) Effects of fine and ultrafine particles on cardiorespiratory symptoms in elderly subjects with coronary heart disease—The ULTRA study. *Am. J. Epidemiol*., **157**, 613–623, DOI: 10.1093/aje/kwg021.

de Mesnard, L. (2013) Pollution models and inverse distance weighting: Some critical remarks. *Comput. and & Geosci.*, **52**, 459–469, http://dx.doi.org/10.1016/j.cageo.2012.11.002

Devarakonda, S., Sevusu, P., Liu, H., Liu, R., Iftode, L., Nath, B. Real-time Air Quality Monitoring Through Mobile Sensing in Metropolitan Areas. Proc. of the *2nd ACM SIGKDD International Workshop on Urban Computing*, August 13, 2013, Chicago, Illinois, p. 1-8, https://doi.org/10.1145/2505821.2505834.

Dockery, D.W., Pope III, A., Xu, X., Spengler, J.D., Ware, J.H., Fay, M.E., Ferris Jr., B.G., Speizer, F.E. (1993) An association between air pollution and mortality in six US cities. *N. Engl. J. Med*., **329**, 1753–1759, DOI: 10.1056/NEJM199312093292401.

Dockery, D.W. (2001) Epidemiologic evidence of cardiovascular effects of particulate air pollution. *Environ. Health Perspect.* **109**, 483–486, DOI: 10.1289/ehp.01109s4483.

Donaire-Gonzalez, D., Valentín, A., van Nunen, E., Curto, A., Rodriguez, A., Fernandez-Nieto, M., Naccarati, A., Tarallo, S., Tsai, M.-Y., Probst-Hensch, N., Gulliver, J., Nieuwenhuijsen, M.J. (2019) ExpoApp: An integrated system to assess multiple personal environmental exposures. *Environ. Int*., **126**, 494–503, https://doi.org/10.1016/j.envint.2019.02.054.

Elbir, T. (2003) Comparison of model predictions with the data of an urban air quality monitoring network in Izmir, Turkey. *Atmos. Environ*., **37**, 2149–2157, 10.1016/S1352-2310(03)00087-6.

Elen, B., Peters, J., Van Poppel, M., Bleux, N., Theunis, J., Reggente, M., Standaert, A. (2013) The Aeroflex: A Bicycle for Mobile Air Quality Measurements. *Sensors*, **13**, 221-240, https://doi.org/10.3390/s130100221.

Elimelech, M.; Gregory, J.; Jia, X.; Williams, R. (1998) *Particle Deposition and Aggregation: Measurement, Modelling and Simulation*, Butterworth-Heinemann, 1998.

Environmental Protection Agency – EPA. (1995) *A: User's Guide for the Industrial Source Complex (ISC3) Dispersion Model*. Volume II.

Environmental Protection Agency – EPA. (2020) *Requirements for preparations, Adoption and Submittal of State Implementation Plans (Guideline on Air Quality Models)*, Proposed Rule - Part II

Environmental Protection Agency – EPA. AERMOD Techical Guide.

Fasogbon, S. K., Oyelami, F. H., Adetimirin, E. O., Ige , E. O. (2019) On blasius plate solution of particle dispersion and deposition in human respiratory track. *Mat. Mod. Eng. Prob.*, **6,** 428–432, doi: 10.18280/mmep.060314.

Ferrero, E., Alessandrini, S., Anderson, B., Tomasi, E., Jimenez, P., Meech, S. (2019) Lagrangian simulation of smoke plume from fire and validation using ground-based lidar and aircraft measurements. *Atmos. Environ*., **213**, 659–674, doi: 10.1016/j.atmosenv.2019.06.049.

Gifford, F. A. (1961) Use of routine meteorological observations for estimating atmospheric dispersion. *Nucl. Saf*., **2,** 47–51.

Giuliano A., Gioiella F., Sofia D., Lotrecchiano N. (2018) A novel methodology and technology to promote the social acceptance of biomass power plants avoiding Nimby Sindrome. *Chem. Eng. Trans*., **67**. https://doi.org/10.3303/CET1867052.

Gustafson, S., Kortanek, K. O., & Sweigart, J. R. (1977) Numerical Optimization Techniques in Air Quality Modeling: Objective Interpolation Formulas for the Spatial Distribution of Pollutant Concentration. *J. Appl. Meteorol. Climatol.*, **16**, 1243-1255, https://doi.org/10.1175/1520-0450(1977)016<1243:NOTIAQ>2.0.CO;2

Hache G., Lemaire ED., Baddour N., (2011) Wearable Mobility Monitoring Using a Multimedia Smartphone Platform. *IEEE Trans. Instrument. Measur*., **60,** 3153 – 3161, DOI: 10.1109/TIM.2011.2122490.

Hao W.-M., Liu M.-H., Crutzen, P. J. (1990) Estimates of annual and regional release of CO2 and other trace gases to the atmosphere from fires in the tropics, based on the FAO statistics for the period 1975-1980. In *Fire in the tropical Biota: Ecosystem processes and Global Challenges* (Goldammer, J. G. Eds.), Springer-Verlag, p. 400-462,

Hart, J.E., Yanosky, J.D., Puett, R.C., Ryan, L., Dockery, D.W., Smith, T.J., Garshick, E.,Laden, F. (2009) Spatial modeling of PM10 and NO2 in the continental United States, 1985–2000. *Environ. Health Perspect.*, **117**, 1690–1696, 10.1289/ehp.0900840.

Hasenfratz D., Saukh O., Walser C., Hueglin C., Fierz M., Arn T., Beutel J., Thiele L. (2015) Deriving high-resolution urban air pollution maps using mobile sensor nodes. *Pervasive Mob. Comput*., **16**, 268-285.

Hertel, O., and Berkowicz, R. (1989) *Modeling pollution from traffic in a street canyon: Evaluation of data and model development*. National Environmental Research Institute, Denmark, p. 77.

Hooyberghs, J., Mensink, C., Dumont, G., Fierens, F. (2006) Spatial interpolation of ambient ozone concentrations from sparse monitoring points in Belgium. *J. Environ. Monit.*, **8**, 1129 – 1135, DOI: 10.1039/b612607n.

OSPM User's guide. Available at: vergina.eng.auth.gr/mech/lat/copert/copert.htm.

Huang, R., Qin, M., Hu, Y., Russell, A. G., Odman, M. T. (2020) Apportioning prescribed fire impacts on PM2.5 among individual fires through dispersion modeling. *Atmos. Environ*., **223**, 17260, doi: 10.1016/j.atmosenv.2020.117260.

Hurley, P. J., Physick, W. L., Luhar, A. K. (2005) TAPM: a practical approach to prognostic meteorological and air pollution modelling. *Environ. Model. Soft.*, **20**, 737–752, doi: 10.1016/j.envsoft.2004.04.006.

Jeftic, L., Kenworthy, J.M., Milliman, J.D., Sestini, G. (1992) *Climatic Change and the Mediterranean: Environmental and Societal Impacts of Climatic Change and Sea-Level Rise in the Mediterranean Region*. Arnold Publisher, London (UK).

Kuberský P., Syrový T., Hamáček A., Nešpůrek S., Syrová L. (2014) Fully Printed Electrochemical NO2 Sensor. *Procedia Eng*., **87**, 1043 – 1046, DOI: 10.1016/j.proeng.2014.11.340.

Karimi, B. and Shokrinezhad, B. (2021) Spatial variation of ambient PM2.5 and PM10 in the industrial city of Arak, Iran: A land-use regression. *Atmos. Pol. Res.*, **12**, 101235, https://doi.org/10.1016/j.apr.2021.101235

Leelőssy, Á., Molnár, F., Izsák, F., Havasi, Á., Lagzi, I., Mészáros, R. (2014) Dispersion modeling of air pollutants in the atmosphere: a review. *Cent. Eur. J. Geosci*., **6**, doi: 10.2478/s13533-012-0188-6.

Lotrecchiano N., Gioiella F., Giuliano A., Sofia D. (2019) Forecasting Model Validation of Particulate Air Pollution by Low Cost Sensors Data. *J. Model. Optim.*, 11, 63-68, 10.32732/jmo.2019.11.2.63.

Lotrecchiano N., Sofia D., Giuliano A., Barletta D., Poletto, M. (2019) Real-time On-road Monitoring network of air quality. *Chem. Eng. Trans*., **74**, 241–246, doi:10.3303/CET1974041.

Lotrecchiano, N., Sofia, D., Giuliano, A., Barletta, D., Poletto, M. (2020) Pollution dispersion from a fire using a Gaussian plume model. *Int. J. Saf. Secur. Eng*, **10**, https://doi.org/10.18280/ijsse.100401

Li X., Ma Y., Wang Y., Liu N., Hong Y. (2017) Temporal and spatial analyses of particulate matter (PM10 and PM2. 5) and its relationship with meteorological parameters over an urban city in northeast China. *Atmos. Res*., **198**, 185-193, DOI: 10.1016/j.atmosres.2017.08.023.

Lim CC., Kim H., Ruzmyn Vilcassima M.J., Thurston GD., Gordon T., Chen LC., Lee K., Heimbinder M., Kim SY. (2019) Mapping urban air

quality using mobile sampling with low-cost sensors and machine learning in Seoul, South Korea. *Environ. Int*., **131**, 105022, DOI: 10.1016/j.envint.2019.105022.

Luhar, A. K., Emmerson, K. M., Reisen, F., Williamson, G. J., Cope, M. E. (2020) Modelling smoke distribution in the vicinity of a large and prolonged fire from an open-cut coal mine. *Atmos. Environ*., **229**, 117471, doi: 10.1016/j.atmosenv.2020.117471.

Luhar, A. K., Mitchell, R. M., Meyer, C. P., Qin, Y., Campbell, S., Gras, J. L., Parry, D. (2008) Biomass burning emissions over northern Australia constrained by aerosol measurements: II-Model validation, and impacts on air quality and radiative forcing. *Atmos. Environ*, **42**, 1647–1664, doi: 10.1016/j.atmosenv.2007.12.040.

Mallia, D., Kochanski, A., Urbanski, S., Lin, J. (2018) Optimizing Smoke and Plume Rise Modeling Approaches at Local Scales. *Atmosphere*, **9**, 166, doi: 10.3390/atmos9050166.

Mandel, J., Beezley, J. D., Kochanski, A. K. (2011) Coupled atmosphere-wildland fire modeling with WRF 3.3 and SFIRE 2011. *Geosci. Model. Dev*., **4**, 591–610, doi: 10.5194/gmd-4-591-2011.

Mareddy, A.R. (2018) Environmental Impact Assessment: Theory and Practice, Butterworth-Heinemann.

Michalakes, J., Dudhia, J., Hart, L., Klemp, J., Middlecoff, J., Skamarock, W. (2001) Development of a next-generation regional weather research and forecast model. In *Developments in Teracomputing*. p.269–276, doi: 10.1142/9789812799685_0024.

Meharg A.A, French M.C. (1995) Heavy metal as marker for assessing the environmental for environmental pollution from chemical warehouse and plastics fires. *Chemosphere*, **30**, 1987-1994.

Mihăiţră A. S., Dupont L., Chery O., Camargo M., Cai C. (2019) Evaluating air quality by combining stationary, smart mobile pollution monitoring and data-driven modelling. *J. Clean. Prod*., **221**, 398-418, https://doi.org/10.1016/j.jclepro.2019.02.179.

Minet, L., Gehr, R., Hatzopoulou, M. (2017) Capturing the sensitivity of landuse regression models to short-term mobile monitoring campaigns using air pollution micro-sensors. *Environ. Pollut*., **230**, 280-290, DOI: 10.1016/j.envpol.2017.06.071.

Mirzaei A., Leonardi S. G., Neri G. (2016) Detection of hazardous volacell organic compounds (VOCs) by metal oxide nanostructures-based gas sensors: A review. *Ceram. Int.*, **42**, 15119–15141, 10.1016/j.ceramint.2016.06.145.

Moussiopoulos, N., Berge, E., Bøhler, T., de Leeuw, F., Grønskei, K.E., Mylona, S., Tombrou, M. (1996) *Ambient air quality, pollutant dispersion and transport models*. European Environment Agency, Copenhagen, Denmark.

OSPM User's guide. Available at: vergina.eng.auth.gr/mech/lat/copert/copert.htm

Parra, M.A., Santiago, J.L., Martín, F., Martilli, A., Santamaría J.M. (2010) A methodology to urban air quality assessment during large time periods of winter using computational fluid dynamic models. *Atmos. Environ.*, **44**, 2089-2097, DOI: 10.1016/j.atmosenv.2010.03.009.

Pasquill, F. (1961) The estimation of the dispersion of windborne material. *Meteor. Mag.*, **90**, 33-49.

Pengpom, N., Vongpradubchai, S., Rattanadecho, P. (2019) Numerical analysis of pollutant concentration dispersion and convective flow in a two-dimensional confluent river model. *Math. Model. Eng. Probl.*, **6**,271–279, doi: 10.18280/mmep.060215.

Pereira, A., Patrìcio, B., Fonte, F., Marques, S., Reis, C.I., Maximiano, M. (2018) Collecting information about air quality using smartphones. *Comput. Sci.*, **138**, 33-40, DOI: 10.1016/j.procs.2018.10.006.

Pirjolaa, L., Parviainena, H., Husseinb, T., Vallia, A., Hämerib, H., Aaaltob, P., Virtanend, A., Keskinend, J., Pakkanene, T.A., Mäkelä, T., Hillamo, R.E. (2004) ''Sniffer''—a novel tool for chasing vehicles and measuring traffic pollutants. *Atmos. Environ.*, **38**, 3625–3635, DOI: 10.1016/j.atmosenv.2004.03.047.

Pope III, C.A., Thun, M.J., Namboodiri, M.M., Dockery, K.S., Evans, J.S., Speizer, F.E., Heath Jr., C.W. (1995) Particulate air pollution as a predictor of mortality in a prospective study of US adults. *Am. J. Respir. Critic. Care Med*, **151**, 669–674, DOI: 10.1164/ajrccm/151.3_Pt_1.669.

Pope III, C.A., Burnett, R.T., Thun, M.J., Calle, E.E., Krewski, D., Ito, K., Thurnston, G.D. (2002) Lung cancer, cardio-pulmonary mortality, and long-term exposure to fine particulate air pollution. *J. Am. Med. Assoc*., **287**, 1132–1141, DOI: 10.1001/jama.287.9.1132.

Price, O. F., Horsey, B., Jiang, N. (2016) Local and regional smoke impacts from prescribed fires. *Nat. Hazards Earth Syst. Sci.*, **16**, 2247–2257, doi: 10.5194/nhess-16-2247-2016.

Raspa G. and Bruno, R. (1994) *Dispense di Geostatistica Applicata*. Capitolo 3 – Geostatistica di base.

Salehi Kahrizsangi H., Sofia D., Barletta D., Poletto M. (2015) Dust generation in vibrated cohesive powders. *Chem. Eng. Trans.*, **43**, 769–774, DOI: 10.3303/CET1543129.

Ryall, D., Derwent, R., Manning, A., Redington, A., Corden, J., Millington,W., Simmonds, P., O'Doherty, S., Carslaw, N., Fuller, G.. (2000) The origin of high particulate concentrations over the United Kingdom. *Atmos. Environ.*, **36**, 1363–1378, DOI: 10.1016/S1352-2310(01)00522-2.

Samet, J.M., Dominici, F., Curriero, F.C., Coursac, I., Zeger, S.L. (2000) Fine particulate air pollution and mortality in 20 US Cities, 1987–1994. *N. Engl. J. Med*., **343**, 1742–1749, DOI: 10.1056/NEJM200012143432401.

Sahanavin N, Prueksasit T, Tantrakarnapa K. (2018) Relationship between PM10 and PM2. 5 levels in high-traffic area determined using path analysis and linear regression. *J. Environ. Sci*., **69**, 105-114, DOI: 10.1016/j.jes.2017.01.017.

Schwartz, J., Laden, F., Zanobetti, A. (2002) The concentration-response relation between PM2.5 and daily deaths. *Environ. Health Persp*., **110**, 1025–1029, doi: 10.1289/ehp.021101025.

Socolofsky, S.A. and Jirka, J.E. (2002) *Environmental Fluid Mechanics Part I: Mass Transfer and Diffusion Engineering* – Lectures, 2nd Edition.

Seinfeld, John (2006) *Atmospheric Chemistry and Physics: From Air Pollution to Climate Change*. Hoboken, New Jersey: John Wiley & Sons.

Shiva Negendra, S.M., Yasa, P.R., Narayana, M.V., Khadirnaikar, S., Rani, P. (2019) Mobile monitoring of air pollution using low cost sensors to visualize spatio-temporal variation of pollutants at urban hotspots. *Sustain. Cities Soc*., **44**, 520-535, DOI: 10.1016/j.scs.2018.10.006.

Sivacoumar, R., Bhanarkar, A.D., Goyal, S.K., Gadkari, S.K., Aggarwal, A.L. (2001) Air pollution modeling for an industrial complex and model performance evaluation. *Environ. Pollut*., **111**, 471–477, https://doi.org/10.1016/S0269-7491(00)00083-X.

Sivacoumar, R. and Thanasekaran, K. (1999) Line source model for vehicular pollution prediction near roadways and model evaluation through statistical analysis. *Environ. Pollut*., **104**, 389–395, https://doi.org/10.1016/S0269-7491(98)00190-0.

Sofia D., Giuliano A., Barletta D. (2013) Techno-Economic Assessment of Co-gasification of Coal-Petcoke and Biomass in IGCC Power Plants. *Chem. Eng. Trans*. **32**, 1231–1236, DOI: 10.3303/CET1332206.

Sofia D., Giuliano A., Gioiella F., Barletta D., Poletto M. (2018a) Modeling of an air quality monitoring network with high space-time resolution. In *Computer Aided Chemical Engineering* (Friedl A., Klemeš J.J., Radl S., Varbanov P.S., Wallek T. Eds), **43**, p. 193-198, DOI: 10.1016/B978-0-444-64235-6.50035-8.

Sofia D., Giuliano A., Gioiella F. (2018b) Air quality monitoring network for tracking pollutants. the case study of Salerno City center. *Chem. Eng. Trans*., **68**, 67-72, DOI: 10.3303/CET1868012.

Sofia D., Lotrecchiano N., Giuliano A., Barletta D., Poletto M. (2019) Optimization of number and location of sampling points of an air quality monitoring network in an urban contest. *Chem. Eng. Trans*. **74**, 277–282, doi:10.3303/CET1974047.

Sofia, D., Gioiella, F., Lotrecchiano, N., Giuliano, A. (2020) Cost-benefit analysis to support decarbonization scenario for 2030: A case study in Italy. *Energy Policy*, **137**, 111137, doi: https://doi.org/10.1016/j.enpol.2019.111137.

Sofia, D., Gioiella, F., Lotrecchiano, N., Giuliano, A. (2020a) Mitigation strategies for reducing air pollution. *Environ. Sci. Pollut. Res*., **27**, 19226–19235, https://doi.org/10.1007/s11356-020-08647-x

Sofia, D., Lotrecchiano, N., Cirillo, D., Villetta, M.L., Sofia, D. (2020b) NO2 Dispersion model of emissions of a 20 kwe biomass gasifier. *Chem. Eng. Trans*., **82**, 451–456, https://doi.org/10.3303/CET2082076

Sofiev, M., Ermakova, T., Vankevich, R. (2012). Evaluation of the smoke-injection height from wild-land fires using remote-sensing data. *Atmos. Chem. Phys.*, **12**, 1995–2006, doi: 10.5194/acp-12-1995-2012.

Steenland, K., Bertazzi, P., Baccarelli, A., Kogevinas, M. (2004) Dioxin revisited: Developments since the 1997 IARC classification of dioxin as a human carcinogen. *Environ. Health Perspec.*, **112**, 1265–1268, doi: 10.1289/ehp.7219.

Stohl, A., Hittenberger, M., Wotawa, G. (1998) Validation of the lagrangian particle dispersion model FLEXPART against large-scale tracer experiment data. *Atmos. Environ.*, **32**, 4245–4264, doi: 10.1016/S1352-2310(98)00184-8.

Sutton, O. G. (1932) A theory of eddy diffusion in the atmosphere. In Proc. of the *Royal Society of London. Series A*, **135**, 143–165, doi: 10.1098/rspa.1932.0025.

Tsujita, W., Yoshino, A., Ishida, H., Moriizumi, T. (2005) Gas sensor network for air-pollution monitoring. *Sens. Actuators B Chem*., **110**, 304–311, https://doi.org/10.1016/j.snb.2005.02.008.

Thakur, J., Thever, P., Gharai, B., Sai, S., and Vnr. Pamaraju. (2019) Enhancement of carbon monoxide concentration in atmosphere due to large scale forest fire of Uttarakhand. *PeerJ*, **7:e6507**, doi: 10.7717/peerj.6507.

Thompson J.E. (2016) Crowd-sourced air quality studies: A review of the literature & portable sensors. *Trends Environ. Anal. Chem.*, **11**, 23–34, http://dx.doi.org/10.1016/j.teac.2016.06.001

Tshehla, C. E., Wright, C. Y. (2019) Spatial and temporal variation of PM10 from industrial point sources in a rural area in limpopo, South Africa. *Int. J. Environ. Health Res*., **16**, doi: 10.3390/ijerph16183455.

Turbelin, G., Singh, S. K., and Issartel, J.-P. (2014) Reconstructing source terms from atmospheric concentration measurements: Optimality analysis of an inversion technique, *J. Adv. Model. Earth Syst*., **6**, 1244– 1255, doi:10.1002/2014MS000385.

Wang J, Song G. (2018) A deep spatial-temporal ensemble model for air quality prediction. *Neurocomputing*, **314**, 198-206, DOI:10.1016/j.neucom.2018.06.049.

Weijers, E.P., Khlystov, A.Y., Kos, G.P.A., Erismana, J.W. (2004) Variability of particulate matter concentrations along roads and motorways determined by a moving measurement unit. *Atmos. Environ.*, **38**, 2993–3002,

10.1016/j.atmosenv.2004.02.045.

Wen Y., Wang H, Larson T., Kelp M., Zhang S., Wu Y., Marshall JD. (2019) On-highway vehicle emission factors, and spatial patterns, based on mobile monitoring and absolute principal component score. *Sci. Total Environ*. **676**, 242–251, https://doi.org/10.1016/j.scitotenv.2019.04.185.

Uda, S. K., Hein, L., Atmoko, D. (2019) Assessing the health impacts of peatland fires: a case study for Central Kalimantan, Indonesia. *Environ. Sci. Pollut. Res.*, **26**, 31315–31327, doi: 10.1007/s11356-019-06264-x.

Van den Bossche J., Peters J., Verwaeren J., Botteldooren D., Theunis J., De Baets B. (2015) Mobile monitoring for mapping spatial variation in urban air quality: development and validation of a methodology based on an extensive dataset. *Atmos. Environ.*, **105**, 148-161, DOI: 10.1016/j.atmosenv.2015.01.017.

Vesilind P.A., Pierce J.J., Weine R.F. (1994) *Environmental Engineering, 3a Ediz*., Butterworth-Heinemann, Boston.

Vismara, R. (1983) *Ecologia applicata all' ingegneria*. Eds. Istituto di Ingegneria Sanitaria.

Vlachogianni, A., Kassomenos, P., Karppinen, A., Karakitsios, S., Kukkonen, J. (2011) Evaluation of a multiple regression model for the forecasting of the concentrations of NOx and PM10 in Athens and Helsinki. *Sci. Total Environ,* **409**, 1559–1571, https://doi.org/10.1016/j.scitotenv.2010.12.040.

Zhu, Q., Liu, Y., Jia, R., Hua, S., Shao, T., Wang, B. 2018. A numerical simulation study on the impact of smoke aerosols from Russian forest fires on the air pollution over Asia. *Atmos. Environ*, **182**, 263–274, doi: 10.1016/j.atmosenv.2018.03.052.