Artificial Intelligence for the Pollution Source Identification

Nicoletta Lotrecchiano[a,b], Diego Barletta[a], Massimo Poletto[a], Daniele Sofia[a,b*]

[a] DIIN-Dipartimento di Ingegneria Industriale, Universitá degli Studi di Salerno, Via Giovanni Paolo II 132, 84084, Fisciano (SA), Italy
[b] Sense Square srl, , Corso Garibaldi 33, 84123, Salerno (SA), 84123, Italy
dsofia@unisa.it

The pollutants dispersion in the environment and the positioning of the emission sources, identified from the measured air quality data, are the main knowledge for defining the environmental status. This study aims to locate the pollution source in the urban environment starting from experimental measurements obtained from the monitoring networks. In detail, the source identification algorithm uses artificial intelligence for source identification (AISI), which was applied to the case study of an industrial paper mill located in a region of southern Italy. In this case study, the air quality monitoring network consists of four smart measuring devices arranged in such a way as to obtain triangular meshes, capable of providing real-time 24/24 h measurements with a resolution of 1 min, information on NO2, PM10, and PM2.5 concentrations, temperature, pressure, relative humidity, wind direction, and intensity. The developed AISI algorithm allows combining the information on the concentrations of pollutants and the wind intensity and direction with their positions to identify the probable pollution source position. Some interesting days were analyzed identifying the pollution sources' locations that were external to the network.

1. Introduction

Air pollution continued to drive a significant burden of premature death and disease with the 307,000 premature deaths during 2019 in the 27EU that were attributed to chronic exposure to fine particulate matter (PM2.5), 40,400 to chronic nitrogen dioxide exposure, and 16,800 to acute ozone exposure (EEA, 2021). The pollutants dispersion in the environment and the identification of the emission sources, defined from the measured air quality data, are the main knowledge for defining the environmental status and for human pollutants absorption definition (Lotrecchiano et al., 2022). Recent studies have tried to define the pollution sources using different modeling approaches. Source apportionment is one of the most widely used and developed methods in the literature. It is defined by Belis (2019) as the technique used to relate emissions from various pollution sources to air pollution concentrations at a given location and period. It can be applied to various pollutants and it is possible to distinguish three principal types of source apportionment results that are potential impacts (Mircea et al. 2019), contributions (Kranenburg et al., 2013), and increments (EEA, 2021). Among the most used is the analysis of the conditional probability function (CPF). In this way the source contribution is calculated, knowing the direction of the wind on the ground and the concentration of origin, using the Positive Matrix Factorization (Han, 2017). Principal component analysis (PCA) is the simplest method and allows, for example, the identification of emissions due to vehicle wear and fuel consumption (Arroyo et al, 2017). By combining PCA analysis with other modeling methodologies, it is possible to implement ensemble models capable of obtaining greater reliability (Li and Yan, 2018). The stochastic approach is based on the use of the potential contribution function of the source (Potential Source Contribution Function - PSCF) to calculate the probability that a source is in a certain position defined by latitude and longitude (Di Talia and Antonioni, 2021). The basis of the dispersion models implemented up to now can be found in simple extrapolations of the Gaussian type, numerical models, multiple linear regressions, and multivariate statistical analysis. However, the studies carried out use air quality measurements from a few concentrated measurement points. By expanding the data collection area, through distributed monitoring systems, it will be possible to implement models for the definition of pollutants dispersion and the search for increasingly precise and fast pollution sources. To increase the effectiveness of

pollution monitoring, it is necessary to identify, with reasonable approximation, the point or points from which a pollutant is emitted into the atmosphere. Only in this way will it be possible to assign responsibility and propose which corrective measures must be taken. In a city or industrial context, positioning the emission source with a certain approximation is a difficult task. This is because there are several potential pollution sources that can also be specific (for example the chimney of an industrial plant) or of a diffuse nature (for example smog from vehicular traffic). An additional complication is represented by the wind and urban morphology which together create pollutant mixing cells making it even more difficult to identify the primary source of pollution. In recent years a lot of attention has been paid to the environment by the industrial sector which aims to maintain a sustainable and eco-compatible activity. For this reason, industrial activities often install air quality monitoring networks in their neighborhood to control their emissions and their impact on the environment. The knowledge not only of the pollution levels but also of the origin of these pollutants can allow the implementation of policies for its mitigation and knowledge by those who live in the proximity of industrial activity on the effect it has on the surrounding environment. In this case study, an algorithm that uses artificial intelligence for source identification (AISI) has been developed and applied to an existing air quality network. The algorithm considers data coming from an air quality network installed in an industrial paper mill and takes into account its position and wind information. AISI doesn't improve the quality of data collected that depends from the measuring device but improves the potentiality of information that can be deduced from them.

2. Materials and methods

2.1 Data used

The monitoring network used in this case study is the one implemented in an industrial paper mill located in Campania municipality in southern Italy. The monitoring site is bordered by a tanning center between the four major Italian poles for the production of leather which extends over an area of approximately 60 km². The monitoring network consists of IoT-based monitoring stations capable of measuring the concentrations of fine particles PM10, PM2.5, PM1, and gases such as NO2. Finally, the measurement devices are completed by the meteorological part represented by the measurements of temperature, humidity, atmospheric pressure, wind direction, and intensity. The accuracy of the sensors is ± 2 μg / m3. The sampling time can be set up to 1 data transmission every 3 minutes, thus obtaining a high temporal resolution of the data that are available 24 / 24h in real-time online through a web portal and a mobile App (Android & IoS). Figure 1 shows the configuration of the installations of the air quality monitoring stations in the industrial paper mill area at points considered interesting for measuring air quality.

Figure 1: Location of the measuring devices in the industrial paper mill area.

The installation points were chosen to take into account the characteristics of the context in which the paper mill is located and in particular, considering the best compromise between the installation possibility and the position of interest, the areas of air stagnation in which pollutants could accumulate and in the way to define whether the emission sources are internal or external to the paper mill site (Sofia et al., 2019). Through this distribution of the monitoring stations, it was possible to obtain the pollution levels throughout the site inside and outside the paper mill. The ultimate aim is to identify any critical issues in the context in which the paper mill is located from the pollution point of view, if, for example, due to vehicular traffic, domestic heating, or other industrial activities as well as the definition of the possible impact that it produces.

2.2 AISI developed algorithm

AISI starts by mapping the territory to be controlled with a series of air quality monitoring stations (assuming a distribution equal to one station every 1-3 km2). The points to be analyzed are arranged, to form adjacent

triangles (mesh) as shown in Figure 1 having one side of the triangle in common, i.e. two sensors, and in each monitoring station air quality data and the meteorological parameters: wind intensity and direction, relative humidity, atmospheric pressure, and pollutants concentration. Data relating to air quality and the weather monitored are analyzed by calculating the average pollution threshold based on the historical data relating to the mesh. To define the mesh, consider the measurement device in which the greatest deviation from the limit threshold identified by Legislative Decree 155/2010 occurs, to identify the position of the atmospheric pollution source.
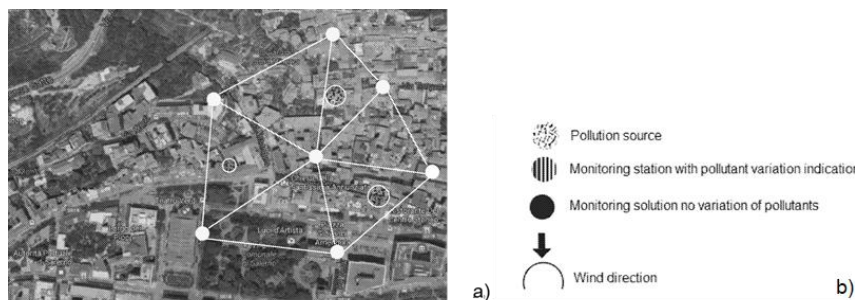

Figure 2: a) Example of mesh definition and b) symbols used in the schematic.

Considering the pollutants measured in the different measuring devices, depending on whether the increase in concentration is identified by none, one, two, or all three of the monitoring stations of a mesh, different situations are identified as illustrated in Figure 3, where the graphic symbols used are summarized in Figure 2b as described in the measuring device's patent (WIPO 2018/225030AI). In particular, Figures 3a and 3b illustrate the case in which only one of the three stations indicates an increase in the concentration of the pollutant. In that case, there are two alternatives (Figure 3a and b).
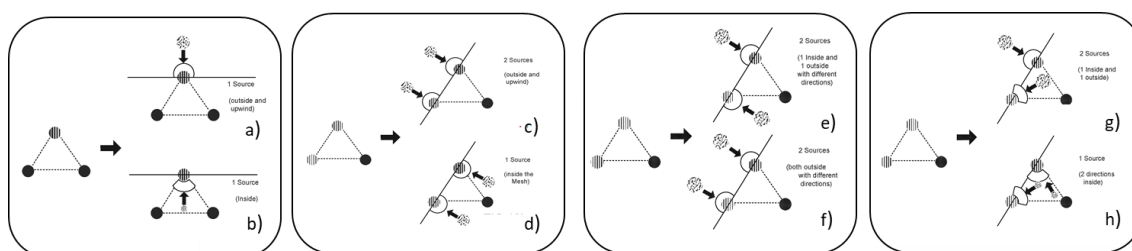

Figure 3: Case in which a-b) only one and c-h) two of the three stations of the same mesh indicate an increase in the concentration of the pollutant.

Figures 3c to 3h illustrate the case in which two of the three monitoring stations of the same mesh indicate an increase in the concentration of the pollutant. In this case, there are six different situations. (1) If the wind of the two stations in which the increase of the concentration comes from the same direction and in particular from the opposite half plane of the mesh (Figure 3c) it can be concluded that the increase in pollution is due to a single source external to the mesh and in particular it is located in the opposite direction to that of the wind. (2) If the wind of the two stations in which the increase of the concentration comes from the same direction and in particular from the semi-plane of the mesh (Figure 3d) it is certain that the increase in pollution is due to a single source inside the mesh. (3) If the wind of the two stations in which the concentrations increase comes from two different directions and in particular one from the half plane of the mesh (Figure 3e) and the other from the half plane opposite to the mesh, it is clear that the increase in pollution is due to two sources of pollution, one external to the mesh and the other internal. (4) If the wind of the two stations in which the concentrations increase comes from two different directions and in particular both from outside the mesh (Figure 3f) then the increase in pollution is due to two sources of pollution, both external to the mesh. (5) If the wind of the two stations in which the increase of the concentrations comes from two different directions and in particular one from the outside of the mesh and one from the inside of the mesh (Figure 3g) then the increase in pollution is due to two sources of pollution, one external to the mesh and one internal. (6) If the wind of the two stations in which the concentrations increase comes from two different directions but both from within the mesh (Figure 3h) then the increase in pollution is due to a single source of pollution inside the mesh.

2.3 Tools used

The AISI algorithm has been developed using the language program Python. The modules used were partly for scientific computing and data manipulation (Numpy) and partly for the manipulation and analysis of geometric objects in the Cartesian plane (Shapely). Since the algorithm uses data stored in a database as JSON, it is necessary to access them by using a request function (Request and JSON module).

3. Results

3.1 Data analysis

The hourly and monthly trends of all the parameters measured by the measuring devices such as PM1, PM2.5, PM10, NO2, temperature, pressure, relative humidity, wind direction, and intensity must be defined. The measured parameters are all linear variables except the wind direction which is a circular variable and as such is appropriately defined in the average calculation. The hourly and monthly trends are plotted on graphs that have been implemented within the system, highlighting any exceedance of the law limits imposed. Figure 5 reports the daily average concentrations measured in point C4 from 1st January 2022 to 31st March 2022 for PM10 (Figure 4a) and PM2.5 (Figure 4b). Point C4 has been chosen as representative of the whole network. From the analysis of Figure 4a, it is clear that during the first three months of the year, at such a point the PM10 concentration only exceeds the law limit twice. In particular, this day is represented by the 1st of January and surely this concentration is influenced mostly by the fireworks that contribute heavily to the particulate concentration. For the PM2.5, during the three months considered, the exceedances were nine (Figure 4b).
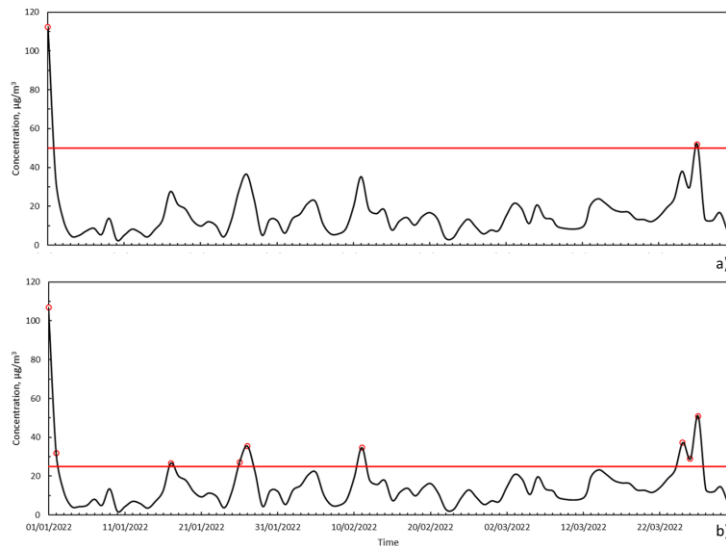

Figure 4: Daily average concentration of a) PM10 and b) PM2.5 measured in the point C3. Concentration are reported in black solid lines, exceedances from the law limit (red solid line) are reported as red dots.

3.2 Algorithm application

The AISI algorithm is applied when the pollutants concentrations exceed the limit threshold. The highest daily concentrations measured from the network have been summarized in Table 1 which also reports the values of the wind intensity and direction. The AISI algorithm has been applied to all days with exceedances but here only two of most interest have been reported (17th January and 27th March 2022). Considering the condition of 17th January, the highest daily concentration has been measured at point C1. On the 27th of March, the highest daily concentration has been measured at point C4. The days' relative wind condition, reported in Table 1, indicates that on the 17th of January, the wind comes from South-West (SW) and goes toward North-East (NE) while on the 27th of March, the wind comes from South-South-East (SSE) and goes toward North-North-West (NNO). So, the AISI algorithm, matching the pollutants levels and the wind direction, individuates the most polluted mesh (red in Figure 5). To define the most polluted mesh, the algorithm considers the highest pollutants' value to be the lowest one.

Table 1: Parameters related to the measuring device with the highest measured particulate concentration exceeding the threshold of the law value according D.Lgs. 155/2010.

Day	Measuring device	PM10 average concentration, $\mu g/m^3$	DailyPM2.5 average concentration, $\mu g/m^3$	DailyWind intensity, km/h	Wind direction
01/01/2022	C4	112.4	107.0	1.3	N
02/01/2022	C4	32.4	35.0	1.8	ESE
17/01/2022	C1	28.0	27.1	14.4	SW
26/01/2022	C3	38.6	38.0	18.0	NW
27/01/2022	C3	40.5	40.0	10.8	NNW
11/02/2022	C3	37.8	37.3	36.0	SW
24/03/2022	C3	30.0	29.4	36.0	NNW
25/03/2022	C3	43.0	42.0	32.4	NNW
26/03/2022	C4	33.3	32.5	21.6	ENE
27/03/2022	C4	52.0	51.0	108.0	SSE

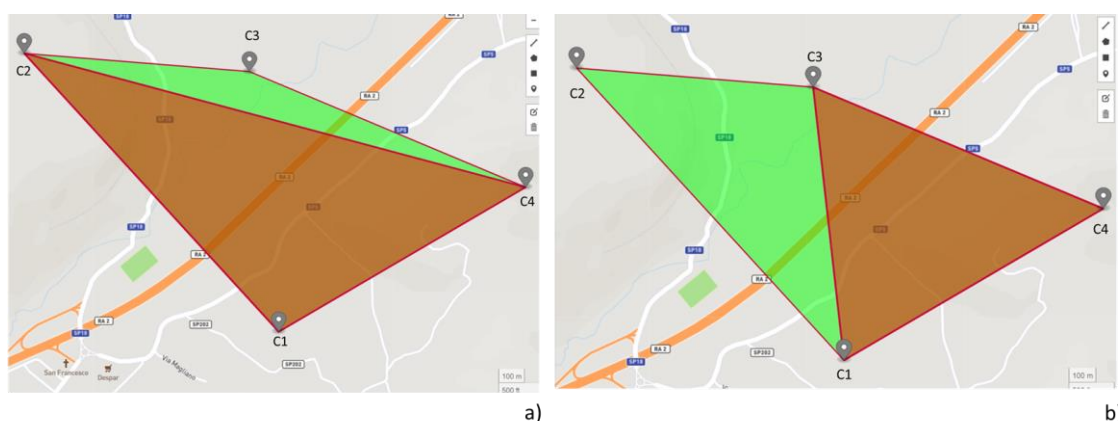
a)   b)

Figure 5: Most polluted mesh identification (red) by the AISI algorithm for a) 17st of January 2022, b) for 27th March 2022. The green area represents the less polluted mesh.

On the first day considered (17th January), the most polluted mesh includes the points C1-C4-C2 while on the second day considered it involves the points C4-C1-C3. Considering the wind direction, as discussed in section 2.2, the source is most probably one and is located at the extern of the mesh on both days. In particular, a deep analysis of March 27th, shows a sudden high increase of pollutants levels starting from 7 PM until 9 PM, both for PM10 and PM2.5 (Figure 6) in C4. The analysis of these accidental events is well described by the air quality monitoring networks as discussed in the literature (Lotrecchiano et al., 2020).
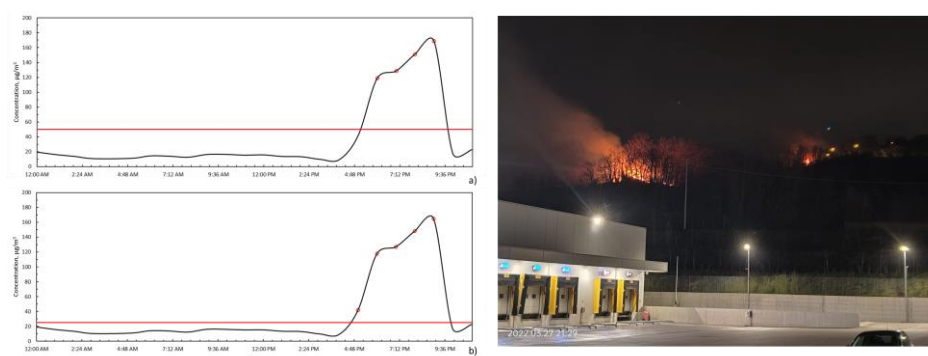

Figure 6: Hourly average concentration of a) PM10 and b) PM2.5 measured in the point C4. Concentration are reported in black solid lines, exceedances from the law limit (red solid line) are reported as red dots. c)Fire occurred near the industrial paper mill.

As the AISI algorithm indicates that the probable location of the pollution source is extern to the mesh, it has been compared with the real happening. When the pollutants overcame the limits, a fire near the paper mill occurred (Figure 6c), so the AISI results match the real case.

4. Conclusions

An algorithm for pollution source identification (AISI) has been implemented and developed using Python language. The developed AISI algorithm define the most polluted mesh of the air quality network considered and locates the pollution source inside or outside the mesh. The algorithm has been applied to the period 1st January 2022-31st March 2022 to an air quality network implemented in an industrial paper mill. Two particular days were reported and, matching the wind direction with the PM10 and PM2.5 concentrations, the possible location of the pollution source (internal or external to the mesh) has been found. Further improvements of AISI will be the inclusion of the wind intensity to define the distance from the source and the knowledge of the urban context to improve the define the dispersion model.

Acknowledgments

References

Belis C.A, Pernigotti D., Pirovano G., Favez O., Jaffrezo J.L., Kuenen J., Denier van Der Gon H., 2020, Evaluation of receptor and chemical transport models for PM10 source apportionment, Atmospheric Environment X 5, 100053, DOI 10.1016/j.aeaoa.2019.100053.

Di Talia V., Antonioni G., 2021, A Model for the Evaluation of Vocs Abatement by Potted Plants in Indoor Environments, Chemical Engineering Transactions, 86, 415-420 DOI:10.3303/CET2186070

European Environmental Agency (EEA), 2011, The application of models under the European Union's Air Quality Directive: A technical reference guide. Copenhagen, European Environment Agency (EEA Technical Report, 10/2011). http://www.eea.europa.eu/publications/fairmode.

Jing Han, 2017, Diffusion law of air pollution in chemical enterprises, Chemical Engineering Transactions, 59, 11531158 DOI:10.3303/CET1759193.

Li Z., Yan X., 2018, Ensemble learning model based on selected diverse principal component analysis models for process monitoring, Journal of Chemometrics, 32, 6, https://doi.org/10.1002/cem.3010.

Lotrecchiano N., Sofia D., Giuliano A., Barletta D., Poletto M., 2020, Pollution dispersion from a fire using a Gaussian plume model, International Journal of Safety and Security Engineering, 10, 431–439, doi:10.18280/ijsse.100401.

Lotrecchiano N., Montano L., Bonapace I.M, Tenore G., 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, https://doi.org/10.3390/pr10071409.

Mircea M., Calori G., Pirovano G., Belis C.A., European guide on air pollution source apportionment for particulate matter with source oriented models and their combined use with receptor models, EUR 30082 EN, Publications Office of the European Union, Luxembourg, 2020, ISBN 978-92-76-10698-2, doi:10.2760/470628, JRC119067.

Kranenburg R., Segers A. J., Hendriks C., Schaap M, 2013, Source apportionment using LOTOS-EUROS: module description and evaluation, Geoscience Model Development, 6, 721–733, https://doi.org/10.5194/gmd-6-721-2013.

World Health Organization (WHO), 2021, Air quality in Europe 2021 - Report no. 15/2021, doi: 10.2800/549289.

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, Chemical Engineering Transactions, 74, 277–282, doi:10.3303/CET1974047.