Article

The Relationship between PM2.5 and PM10 in Central Italy: Application of Machine Learning Model to Segregate Anthropogenic from Natural Sources

Carlo Colangeli [1,2], Sergio Palermi [1], Sebastiano Bianco [1], Eleonora Aruffo [3,4], Piero Chiacchiaretta [2,4,*] and Piero Di Carlo [3,4]

1   Regional Agency for Environmental Protection of Abruzzo (Arta Abruzzo), 66100 Chieti, Italy; c.colangeli@artaabruzzo.it (C.C.); s.palermi@artaabruzzo.it (S.P.); s.bianco@artaabruzzo.it (S.B.)
2   Department of Psychological, Health and Territory Science, University of G. D'Annunzio of Chieti-Pescara, 66100 Chieti, Italy
3   Department of Advanced Technologies in Medicine and Dentistry, University G. D'Annunzio of Chieti-Pescara, 66100 Chieti, Italy; eleonora.aruffo@unich.it (E.A.); piero.dicarlo@unich.it (P.D.C.)
4   Center for Advanced Studies and Technology (CAST), University G. D'Annunzio of Chieti-Pescara, 66100 Chieti, Italy
*   Correspondence: piero.chiacchiaretta@unich.it

Abstract: Particular Matter (PM) data are the most used for the assessment of air quality, but it is also useful to monitor VOC and CO. The health impact of PM increases with decreasing aerodynamic dimensions, therefore most of the monitoring is aimed at PM10 (fraction of PM with aerodynamic dimensions smaller than 10 μm) and PM2.5 (fraction with aerodynamic dimensions lower than 2.5 μm). Generally, anthropogenic emissions contribute mainly to PM2.5 levels, whereas natural sources can largely affect PM10 concentrations. PM2.5/PM10 ratio can be used as a proxy of the origin (anthropogenic vs natural) of the PM, providing a useful indication about the main sources of PM that characterizes a specific geographical or urban setting. This paper presents the results of the analysis of continuous measurements of PM10 and PM2.5 concentrations at eight stations of the regional air quality monitoring network in Abruzzo (Central Italy), in the period 2017–2018. The application of models based on machine learning technique shows that PM2.5/PM10 ratio can be used to classify PM emissions and to know the nature of the emission source (natural and anthropogenic), under determinate conditions, and properly taking into account the meteorological parameters.

Keywords: PM10; PM2.5; machine learning; aerosol; neural network; forecast; air pollution; emission source

1. Introduction

The concentration of most air pollutants is influenced by weather conditions. For example, usually pollutant concentrations decrease when meteorological parameters such as wind speed, precipitation, and relative humidity increase, due to more efficient dilution and dry deposition [1]. Increasing atmospheric pressure is usually positively correlated with pollutant accumulation. High temperatures combined with higher humidity can contribute to the increase of PM10 and PM2.5 [2]. Effective control and reduction of air pollution requires good knowledge of the impacts of meteorological parameters on PM10 and PM2.5 concentrations: the combination of meteorological conditions with the dynamic of the boundary layer height results in typical diurnal and seasonal changes of the concentrations of PM10 and PM2.5, determining higher values at night and in the winter [3].

In the last years, different modelling approaches have been developed to analyze and predict the evolution of PM10 and PM2.5. Some of them are based on machine learning techniques [4–8].

A comparison of the performance of different regression models and Artificial Neural Networks (ANN) models based on different architectures shows that, in terms of agreement between measured and forecasted PM10 and PM2.5, the Elman Recurrent approach gives better results compared with ANN without recurrent approach and with multiple linear regression model (MLR) [6,8]. ANNs combined with clustering algorithms is another approach that has shown better forecast capacity than those based on a simple ANN or MLR [9]. Artificial neural network (ANN) approaches are commonly used in many applications of atmospheric science [7,10–12]. In terms of PM2.5 forecast, which is usually less frequently observed than PM10, the inclusion of PM10 in the input variables significantly improves the results of the forecast [4]. Model results are strongly influenced by the distribution of PM10 data: the use of PM10 data with uniform distribution can lead to a more appropriate prediction at high concentrations, but the accuracy of prediction comes down to low PM10 concentrations. To overcome this problem, two training datasets with different distributions could be combined in two prediction models from PM10, obtaining a model suitable for predicting low to high PM10 concentrations [13]. One more aspect of the model architecture is the algorithm that can be linear or non-linear. Different intercomparisons of the results of simulation of models based on different algorithms show that the non-linear algorithm is better performing than the linear algorithm in terms of lower values error, higher precision, and robustness [14].

Since 2017, the air quality in the Abruzzo region (central Italy) has been monitored by a network made up of 16 fixed stations, equipped with a total of over 60 automatic analyzers and managed by the Regional Agency for Protection of the Environment (ARTA Abruzzo, Italy, Pescara), which also validates and publishes the data on the sira.artaabruzzo.it website (last access on 13 December 2021).

In this work, we carried out an analysis of two years (2017–2018) of continuous measurements of PM10 and PM2.5 and ancillary parameters in eight stations of this network. The main goal is to explore how the PM2.5/PM10 ratio can provide information on the origin of PM. In particular, we have analyzed daily and seasonal trends of this ratio at different locations, together with meteorological parameters, relying on machine learning model simulations to segregate the anthropogenic sources from those of natural PM, as indicated in the flowchart of the study (Figure 1).
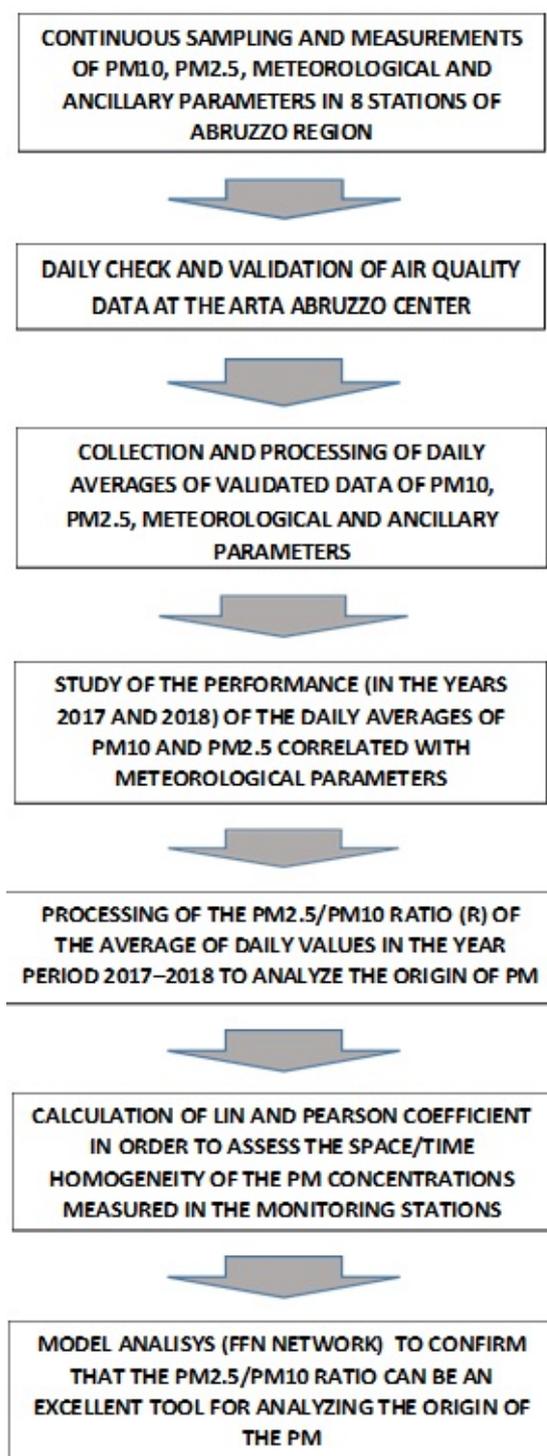
Figure 1. Flowchart of the study steps.

2. Methodology

2.1. Study Area

Data used in this study have been collected in Abruzzo, a region of central Italy characterized by the presence of the highest peak of the Apennines mountains range: the Gran Sasso massif (2912 m asl), located in the west/north-west (about 60 km from the Adriatic coast), and Mount Maiella (2793 m asl) which is located in the south-west (about 35 km from the Adriatic coast) (Figure 2). These orographic features strongly influence the

meteorology of the region, which is subject to meteorological processes such as sea and mountain breezes and convective processes [2].

Figure 2. Air quality monitoring stations in the coastal area of Abruzzo (central Italy). List of the names of the monitoring stations: Sacco (SA), Firenze (FI), Montesilvano (MO), Scuola Antonelli (CH) Francavilla (FR), Amiternum (AQ), S. Gregorio (SG), Gammanara (GA), Porta Reale (PR), Cepagatti (CE) Villa Caldari (OR), Atessa (AT), Castel di Sangro (CS), Arischia (AR), Parco N zionale Maiella (PNM).

The coastal area of Abruzzo is characterized by high humidity all year round, warm summer with temperatures up to 30 °C and a mild winter, typical of the Mediterranean climate. PM10 and PM2.5 concentrations show the typical annual cycle with higher concentrations during the winter and lower concentration during the summer. The most populous city along the Adriatic coast is Pescara (42°27′51.4″ N, 14°12′51.08″ E; located at the estuary of the Aterno-Pescara river) with approximately 120,000 residents and 300,000 in the surrounding metropolitan area. Moreover, Pescara has an international airport (Abruzzo airport) within the urban area and the busiest ports of the area. The monitoring stations, the object of this study, that simultaneously measure the concentration of PM10 and PM2.5, are the following: Teatro D'Annunzio (TH), Via Firenze (FI), Montesilvano (MO), Scuola

Antonelli (CH), Francavilla (FR), Amiternum (AQ), Villa Caldari (OR), and Castel di Sangro (CS) (Table 1).

Table 1. Summary table of the fixed stations of the Abruzzo regional network for monitoring air quality, with details of the measured parameters.

Area	Municipality	Station Name	Station ID	Latitude	Longitude	Type	PM10	PM2.5
Agglomerate Chieti-Pescara (AGG)	Pescara	T.D'annunzio	TH	N 4,700,733 m	E 437,102 m	UB	X	X
	Pescara	Via Sacco	SA	N 4,700,366 m	E 434,150 m	UB	X	
	Pescara	Via Firenze	FI	N 4,702,020 m	E 435,376 m	UT	X	X
	Montesilvano	Montesilvano	MO	N 4,707,801 m	E 430,126 m	UT	X	X
	Chieti Scalo	Scuola Antonelli	CH	N 4,688,783 m	E 429,050 m	UB	X	X
	Francavilla al Mare	Francavilla	FR	N 4,697,015 m	E 429,050 m	UB	X	X
Greater Anthropic Pressure (MAXP)	L'aquila	Amiternum	AQ	N 4,691,713 m	E 366,938 m	UB	X	X
	L'aquila	S. Gregorio	SG	N 4,687,738 m	E 375,604 m	SB		
	Teramo	Gammanara	GA	N 4,724,660 m	E 395,690 m	UB		X
	Teramo	Porta Reale	PR	N 4,723,748 m	E 394,297 m	UT	X	
	Cepagatti	ASL Cepagatti	CE	N 460,147 m	E 423,332 m	RB		
	Ortona	Villa Caldari	OR	N 4,682,708 m	E 446,950 m	SB	X	X
	Atessa	Atessa	AT	N 4,665,673 m	E 453,840 m	I	X	
Lower Anthropic Pressure (minp)	Castel di Sagro	Castel di Sangro	CS	N 4,625,609 m	E 425,526 m	SB	X	X
	Arischia	Arischia	AR	N 4,697,123 m	E 364,389 m	RB		
	S. Eufemia A Majella	Parco Nazionale Maiella	PNM	N 4,663,534 m	E 419,701 m	RB		

The monitoring stations TH, FI, MO, and FR are located along the coastline: TH and FR are urban background stations, while FI and MO are urban traffic stations. The monitoring stations CH and OR are located in the immediate hinterland, whereas AQ and CS are in the Apennines hinterland: CH and AQ are urban background stations, while OR and CS are suburban background stations (Figure 2 and Table 1).

2.2. Model Analysis

In recent years, ANNs that use multiple stages of nonlinear computation (also known as "deep learning") have been able to obtain outstanding performance on an array of complex tasks ranging from visual object recognition to natural language processing. However, it has been found in literature that most of the available tutorials on ANNs are either dense with formal details and contain little information about implementation or any examples, while others skip a lot of the mathematical detail and provide implementations that seem to come from thin air. This post aims to give a more complete overview of ANNs, including (varying degrees of) the math behind ANNs, how ANNs are implemented In code, and finally some examples that point out the strengths and weaknesses of ANNs.

The simplest ANN takes a set of observed inputs, multiplies each of them by their own associated weight, and sums the weighted values to form a pre-activation. Oftentimes there is also a bias that is tied to an input that is always +1 included in the preactivation calculation. The network then transforms the pre-activation, using a nonlinear activation function to output a final activation.

There are many options available for the form of the activation function, and the choice generally depends on the task we would like the network to perform.

For instance, if the activation function is the identity function which outputs continuous values, then the network implements a linear model akin to those used in standard linear regression. Another choice for the activation function is the logistic sigmoid. When the network outputs use the logistic sigmoid activation function, the network implements linear binary classification. Single-layered neural networks used for classification are often referred to as "perceptrons", a name given to them when they were first developed in the late 1950s.

Artificial neural networks (ANN) are based on training and learning methods, thus creating predictive schemes, including the identification and classification of processes. The basic architecture of an ANN includes three parts [15,16]: the input layer (containing neurons or nodes), one or more hidden layers (where other neurons are present), and the output layer (with the respective output neurons). Given the i-th input neurons identified as a vector of X of N elements, in the first step there is the multiplication of each element of the input vector X by the corresponding weight W and then added together (X × W). To produce the last net input, a bias b is added, which is the argument of the function of the output term value generated for the i-th neuron. The training method consists of fine-tuning the bias and weight values, which are randomly set until the network performance is initialized. A hidden single-layer feedforward neural network (FNN), with MLP (Multi-layer Perceptron) [17,18] and different numbers of input neurons and one output neuron, was chosen for these simulations. The information in this type of network (FNN) has the characteristic of propagating only forward through the nodes of the network itself, contrary to recurrent neural networks in which there are feedback connections between the levels [5].

2.3. Sampling and Data Analysis

As previously outlined, PM10 and PM2.5 are detected simultaneously in eight stations of the regional network, five of which are located in the urban agglomeration Pescara-Chieti (AGG), two are in the area with higher anthropogenic pressure (MAXP), and one is in the area with lower anthropogenic pressure (MINP), as shown in Table 1. As expected, the average of the values detected in the two-year period 2017–2018 shows that the highest concentrations are found in the AGG and in the MAXP, whereas the CS station, located in the MINP, shows the lowest values (Figures 3 and 4). There are also evident seasonal variations, with the spring–summer period (from April to September) characterized by significantly lower average values. The only exception concerns the PM10 detected in the CS station, probably due to the prevalence of the natural component. Seasonal variations are more pronounced for PM2.5, which is more affected by the anthropogenic contribution (Tables 2 and 3).

The average values of PM10 are all below the annual limit of 40 $\mu g \cdot m^{-3}$ established by Directive 2008/50/EC of the European Parliament (Strasbourg, France) and of the Council of 21 May 2008 (relating to ambient air quality and cleaner air in Europe), even at the stations where the anthropogenic emissions are higher. Additionally, for PM2.5 there is full compliance with the limit value of 25 $\mu g \cdot m^{-3}$, with average values around 16 $\mu g \cdot m^{-3}$ (denoting substantial homogeneity within the AGG, except for FR), resulting far from those measured in 2017 in the major metropolitan cities of the Po basin (Turin: 33 $\mu g \cdot m^{-3}$, Milan: 30 $\mu g \cdot m^{-3}$, Venice: 29 $\mu g \cdot m^{-3}$) and rather similar to those measured in the large central-southern agglomerations such as Rome (17 $\mu g \cdot m^{-3}$), Florence (16 $\mu g \cdot m^{-3}$), and Bari (15 $\mu g\, m^{-3}$) [19]. The boxplots in Figures 3 and 4 synthesize the data distributions observed in each station. It is interesting to note how the FR station, located in a peripheral area of the AGG, shows a significantly different distribution from the other four AGG stations, resulting in being more similar to the OR station. The other four AGG stations show rather similar average values, with the background stations (TH and CH) characterized by greater variability than the station more impacted by traffic emissions (FI and MO). It is interesting to observe that in the summer semester, when the meteorological conditions favor the dilution of pollutants in the planetary boundary layer, the average PM2.5 level in OR is quite close to that typical of AGG stations; this homogeneity could be related

to the spatial proximity of OR to the AGG (see Figure 2), while the remaining stations (AQ and CS), located in the Apennine hinterland, show significantly lower values. On the other hand, in the winter semester the differences between the various stations are much more marked. The average annual and winter PM2.5/PM10 ratio (Figure 5) do not show marked differences between the various stations, with values ranging between 0.60 and 0.70 and 0.67 and 0.75, respectively. On the contrary, as indicated in the summary table of the average of daily values of the PM2.5/PM10 ratio in the eight stations of the network in the two-year period 2017–2018 (Table 4), in summer the stations located at distances from the sea of less than 1 km (TH, FI, MO, FR) show a PM2.5/PM10 ratio smaller than the remaining stations, with values ranging between ~0.6 and ~0.75. In detail, the TH station, located just 250 m from the coastline, in June–August shows a peculiar PM2.5/PM10 value compared with other stations, probably because it is significantly affected by marine aerosol. The MO, FI, and FR stations also suffer this influence, albeit to a lesser extent, while in the innermost stations, including the OR (located about 7 km from the coast), the ratio remains above 0.6 even in the summer months. The role of marine aerosol in raising PM10 levels in summer in the stations closest to the coast is obviously connected to sea breezes (blowing from north-east), as clearly highlighted by the wind-rose for Pescara, shown in Figure 6b. In wintertime, on the other hand, westerly winds dominate, i.e., blowing from the hinterland (Figure 6a). Of particular importance are the SSW and SW winds, aligned with the axis of the Pescara river valley, typical of anticyclonic periods during which the conditions of atmospheric stability favor thermal inversion and a general worsening of air quality. To confirm this, a polar plot for station TH (Figure 7a,b) has been analyzed: it shows the evident influence of SW winds when the wind speed is less than 1 m s^{-1}, determining the highest concentrations of PM2.5 (Figure 7a). For this urban background station, in fact, the anthropogenic sources of pollution (road networks and industrial sites) are located in the urban area that develops towards the south-west, in the hinterland along the Pescara valley. On the other hand, the analogue polar plot for PM10 (Figure 8b) has significantly different patterns compared with that of PM2.5: the SW winds loses relevance, while the role of NW and NE winds (sea breeze) is dominant in determining high levels of the pollutant, mainly of natural origin (primarily marine aerosol). The comparison between the polar plots of the PM2.5/PM10 ratio at TH and CH (Figure 8a,b) confirms that the SSW-WNW winds determine the highest values of the PM2.5/PM10 ratio (Figure 8a), causing advection of PM2.5 of anthropogenic origin.

In Figure 8b we show the polar plot for the CH station, which highlights high PM2.5/PM10 values related to weak wind speed from the western quadrants, where, with respect to the measurement site, the major anthropogenic sources of pollution are located. The scatterplot in Figure 9 compares the distribution of PM2.5 measured at TH and CH (urban background station), both in summer and winter semesters, whereas the same comparison in Figure 10 concerns the traffic station FI and MO. In Table 5 the values of the Pearson coefficient ρ_{xy} are reported for all the pairs of measurements of PM2.5 for the 5 stations belonging to the AGG plus OR, both for the summer and winter semester. This analysis shows that urban background stations belonging to AGG (TH, CH, FR) are all correlated each other in winter; it is also important to note that the highest correlation value is recorded in winter between TH and FI, with the second station being a traffic station but located only 2 km far from the first. The two traffic stations (FI-MO) show high correlation values in both seasons, with the highest value being recorded in summer. The OR station, as a peripheral station not belonging to the AGG, correlates with all the others to a lesser extent. An explanation may lie in the fact that in winter the atmospheric stability often occurs, favoring the accumulation of PM2.5 concentrations and their spatial homogenization over large geographical areas, for which the urban background stations tend to record values similar to traffic stations, which are located near the emission sources. On the contrary, the weather conditions typical of summer, in which phenomena of greater turbulence and greater height of the boundary layer occur, favor the dispersion of pollutants, increasing the spatial concentration gradients and consequently reducing the degree

of correlation between stations. On the contrary, the traffic stations show a slight increase of the concentrations of PM, since it is near the emission source.
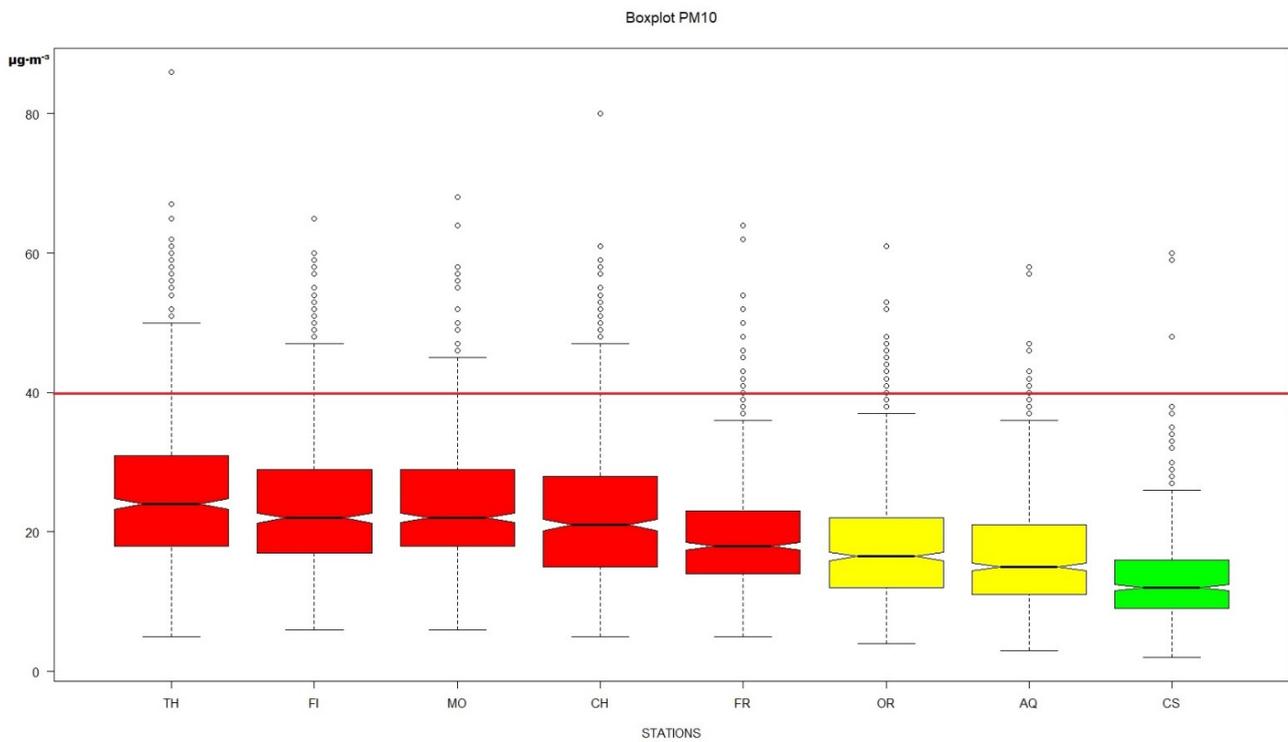

Figure 3. Boxplot of PM10 distributions in $\mu g \cdot m^{-3}$ in the eight stations of the network, two-year period 2017–2018: in red are the stations in AGG, in yellow are those in MAXP, and in green is the station in MINP. The red line indicates the annual EU limit for PM10.
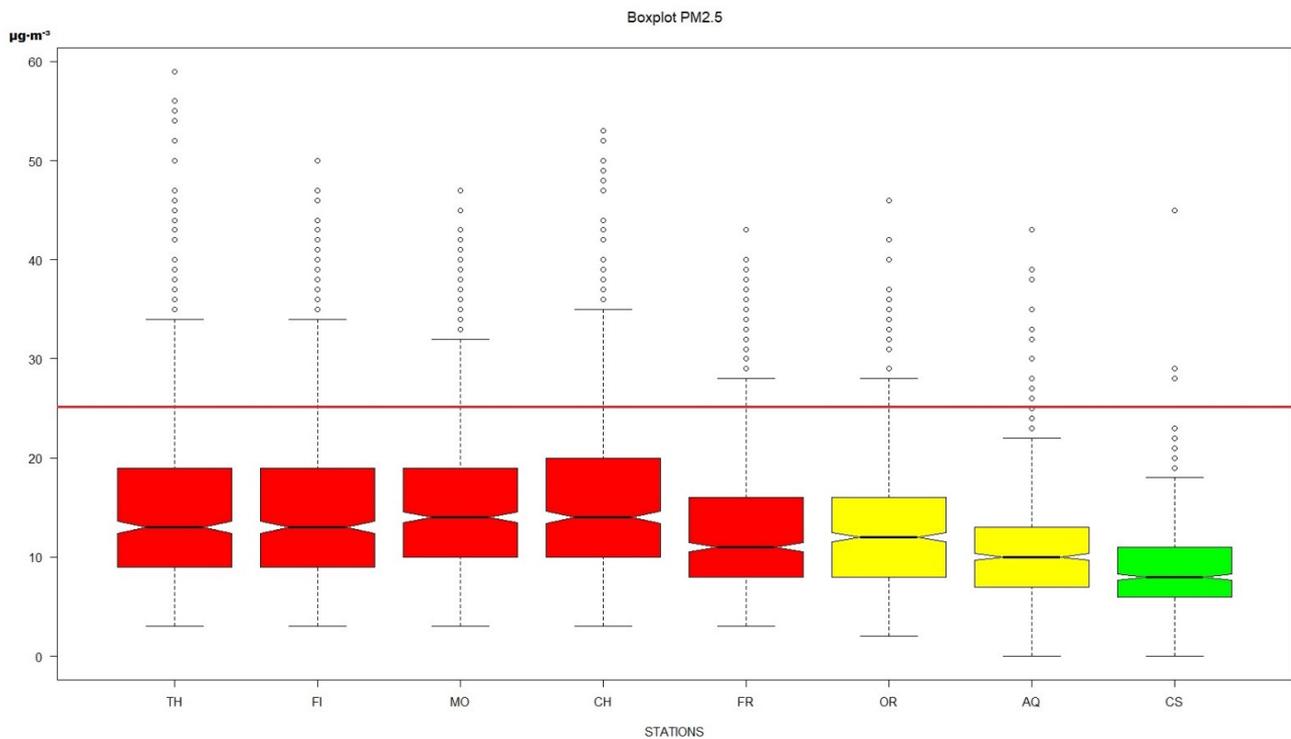

Figure 4. Boxplot of PM2.5 distributions in $\mu g \cdot m^{-3}$, 2017–2018: in red are the stations in AGG, in yellow are those in MAXP, and in green is the station in MINP. The red line indicates the annual EU limit for PM2.5.

Table 2. Summary statistics for PM2.5 (expressed in µg·m^{-3}) measured in the eight stations of the
network in the two-year period 2017–2018, for summer semesters (from April to September), winter
semester (from October to March), and for the whole year.

Period	Statistic	TH	FI	MO	CH	FR	OR	AQ	CS
winter sem.	mean	20.2	19.7	20.0	20.5	16.2	14.5	12.9	9.2
	median	18	18	18	19	14	13	12	9
	st.dev.	11.8	10.7	9.4	10.9	9.1	8.1	7.1	4.6
	min	3	3	3	3	3	2	2	1
	max	59	50	47	53	43	46	43	29
summer sem.	mean	11.5	11.9	11.9	11.9	10.3	11.0	9.3	8.5
	median	11	12	12	12	10	11	9	8
	st.dev.	4.1	4.2	3.8	4.4	4.1	4.4	4.0	3.9
	min	3	4	4	4	3	2	2	2
	max	31	30	27	29	28	25	45	30
year	mean	15.9	15.8	15.9	16.3	13.2	12.7	11.1	8.8
	median	13	13	14	14	11	12	10	8
	st.dev.	9.9	9.0	8.2	9.4	7.6	6.7	6.0	4.3
	min	3	3	3	3	3	2	2	1
	max	59	50	47	53	43	46	43	45
	data av%	95.1	97.0	96.3	91.1	96.7	96.0	97.9	91.9

Table 3. Summary statistics for PM10 (expressed in µg·m^{-3}) measured in the eight stations of the
network in the two-year period 2017–2018, for summer semester (from April to September), winter
semester (from October to March), and for the whole year.

| Period	Statistic	TH	FI	MO	CH	FR	OR	AQ	CS
winter sem.	mean	28.1	28.4	28.2	26.8	21.7	19.0	18.4	12.7
	median	27	27	27	26	20	17	18	12
	st.dev.	12.6	12.4	11.2	13.0	10.0	9.6	9.0	6.2
	min	6	6	7	5	6	4	3	2
	max	65	65	68	80	62	53	47	38
summer sem.	mean	23.7	20.8	20.3	19.1	17.7	17.0	15.3	13.4
	median	23	20	20	18	17	16	14	12
	st.dev.	8.7	6.5	6.0	6.8	6.3	7.0	6.9	6.8
	min	5	6	6	7	5	4	4	4
	max	86	48	46	53	64	61	58	60
year	mean	25.9	24.6	24.3	23.1	19.7	18.0	16.8	13.1
	median	24	22	22	21	18	16.5	13	12
	st.dev.	10.8	10.4	9.6	10.9	8.4	8.3	8.0	6.5
	min	5	6	6	5	5	4	3	2
	max	86	65	68	80	64	61	58	60
	data av%	94.8	96.9	96.9	92.5	97.1	96.0	98.1	91.5
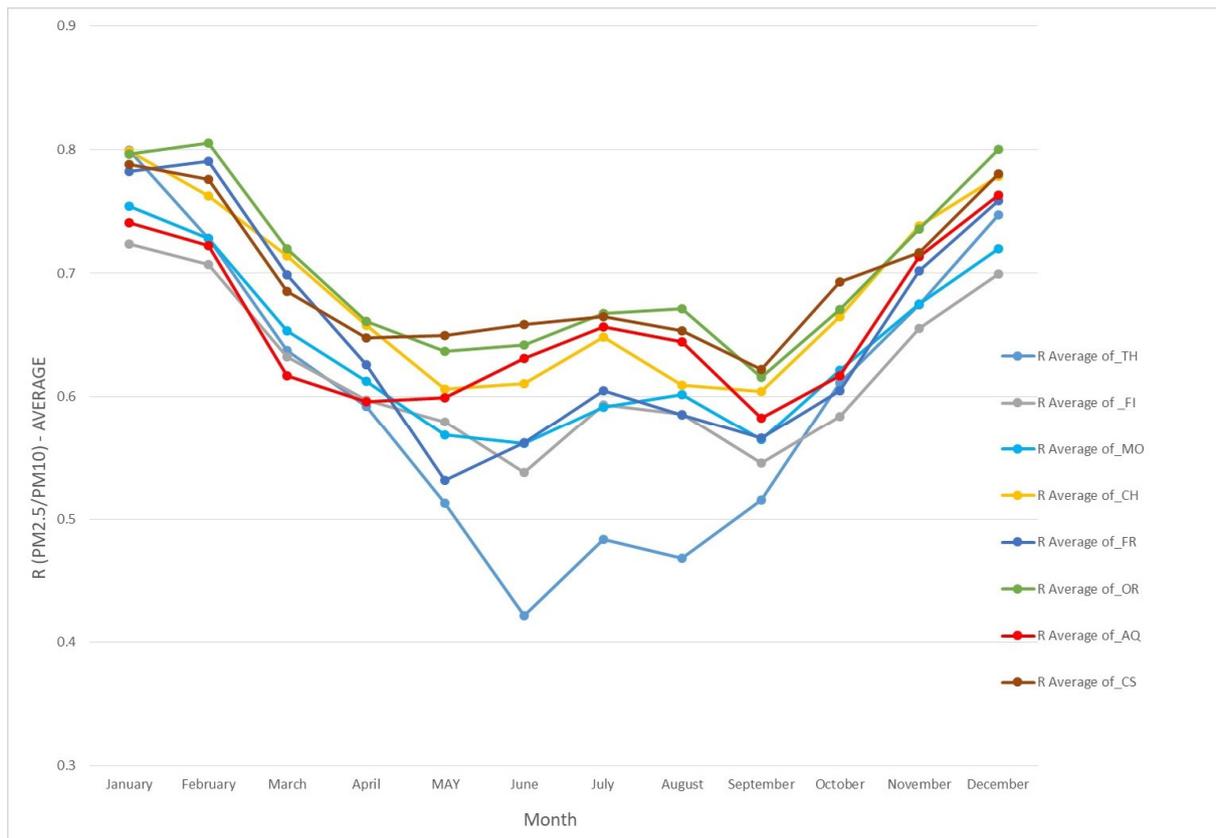
Figure 5. Graph of the monthly average values of R (PM2.5/PM10 ratio) in the eight stations of the regional network in the two-year period 2017–2018.
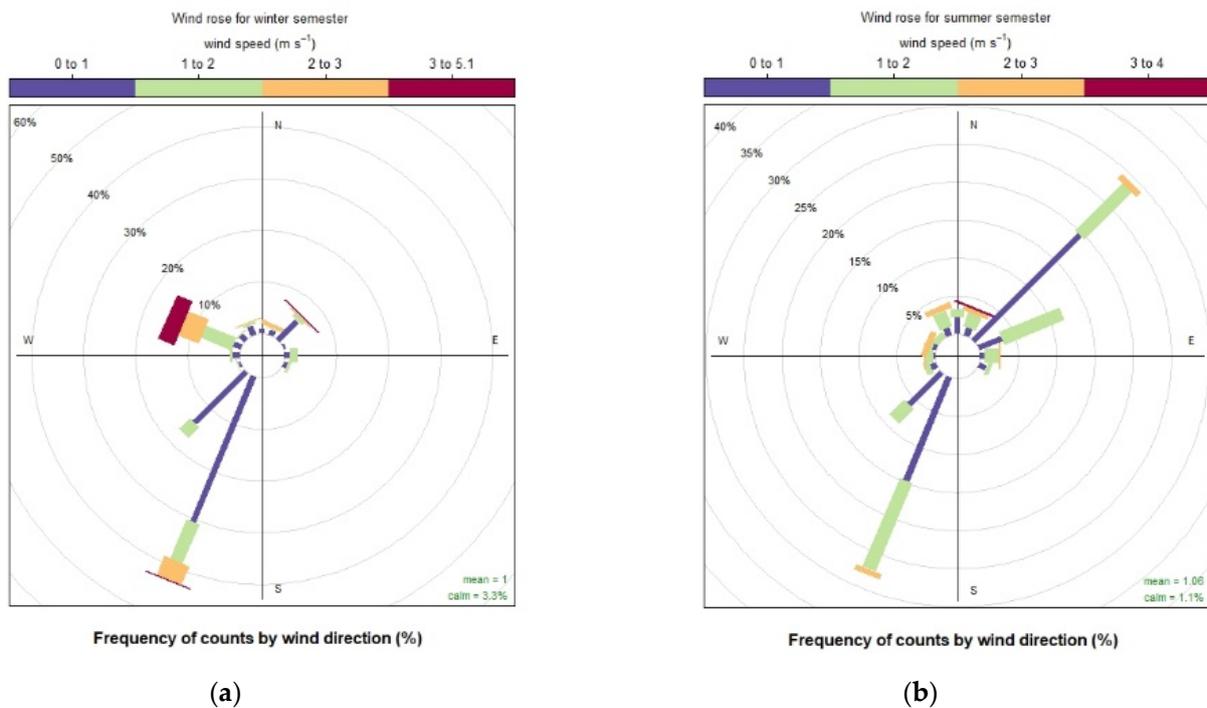

(**a**)

(**b**)

Figure 6. (**a**) Seasonal Windrose (winter semester) for the Pescara weather station, 2017–2018 two-year period. (**b**) Seasonal Windrose (summer semester) for the Pescara weather station, 2017–2018 two-year period.
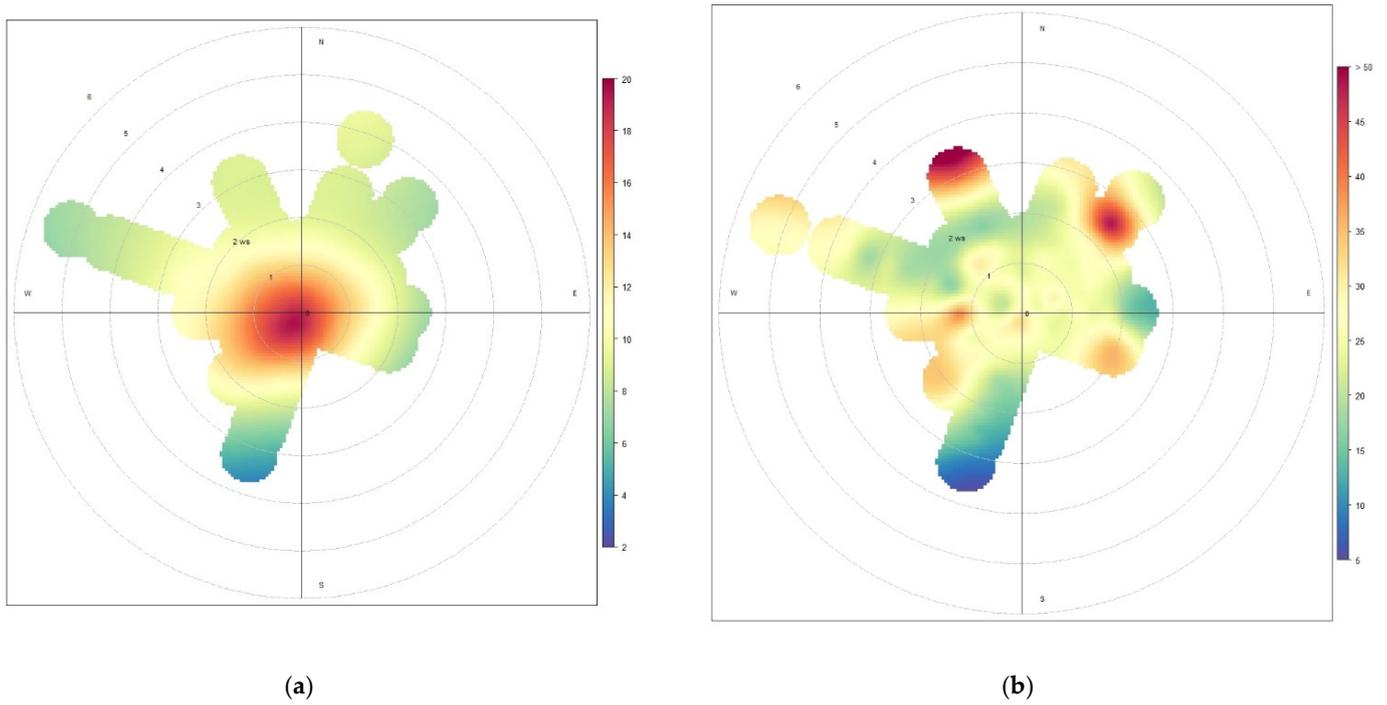
(**a**)

(**b**)

Figure 7. (**a**) Polarplot for PM2.5, TH station, 2017–2018 biennium. (**b**) Polarplot for PM10, TH station, 2017–2018 biennium.
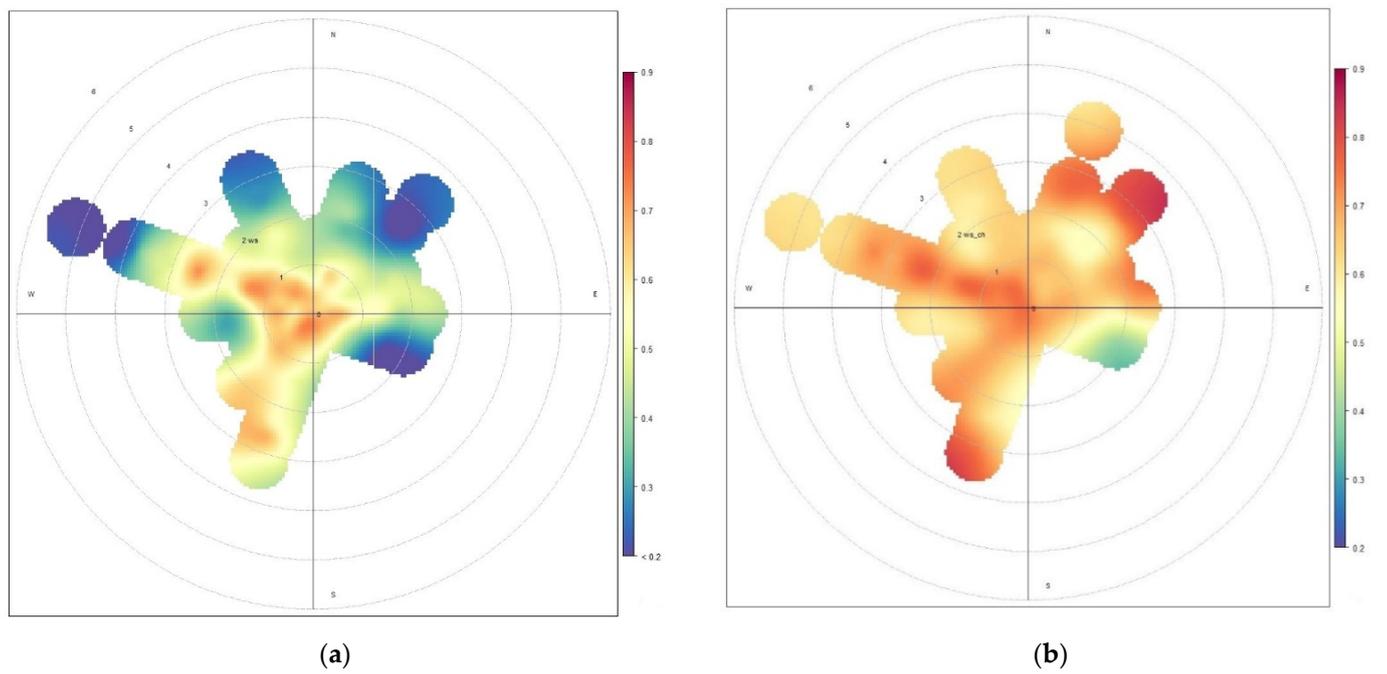

(**a**)

(**b**)

Figure 8. (**a**) Polarplot of the PM2.5/PM10 ratio at TH, 2017–2018. (**b**) Polarplot of the PM2.5/PM10 ratio at CH, 2017–2018.
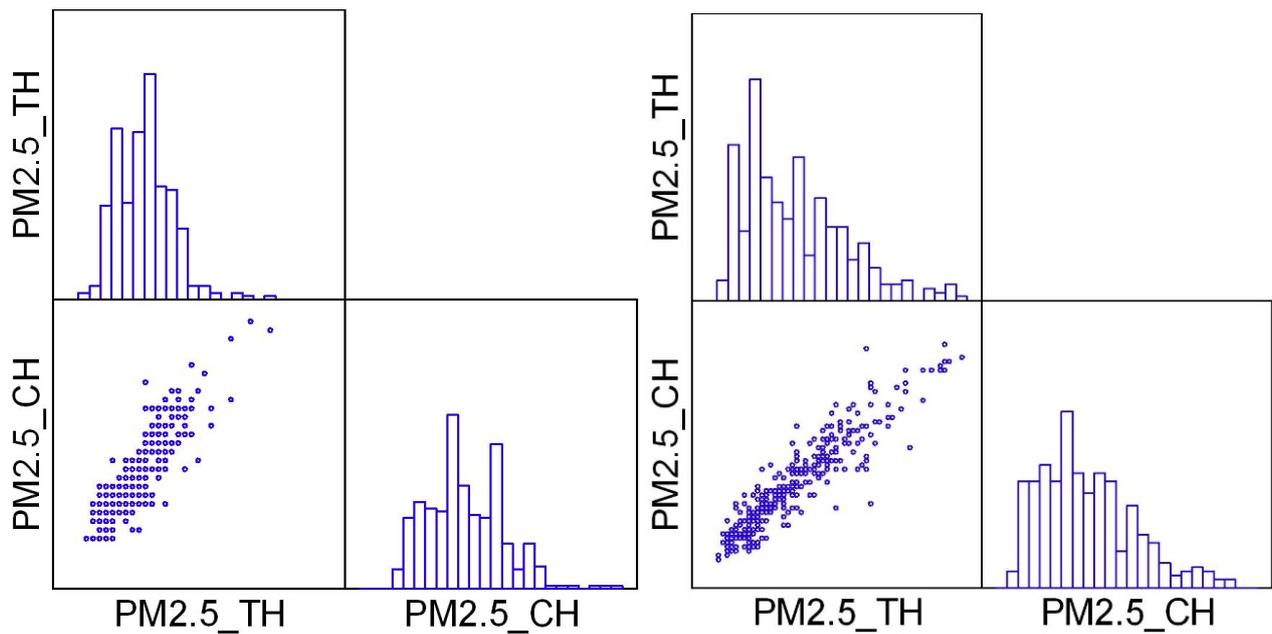
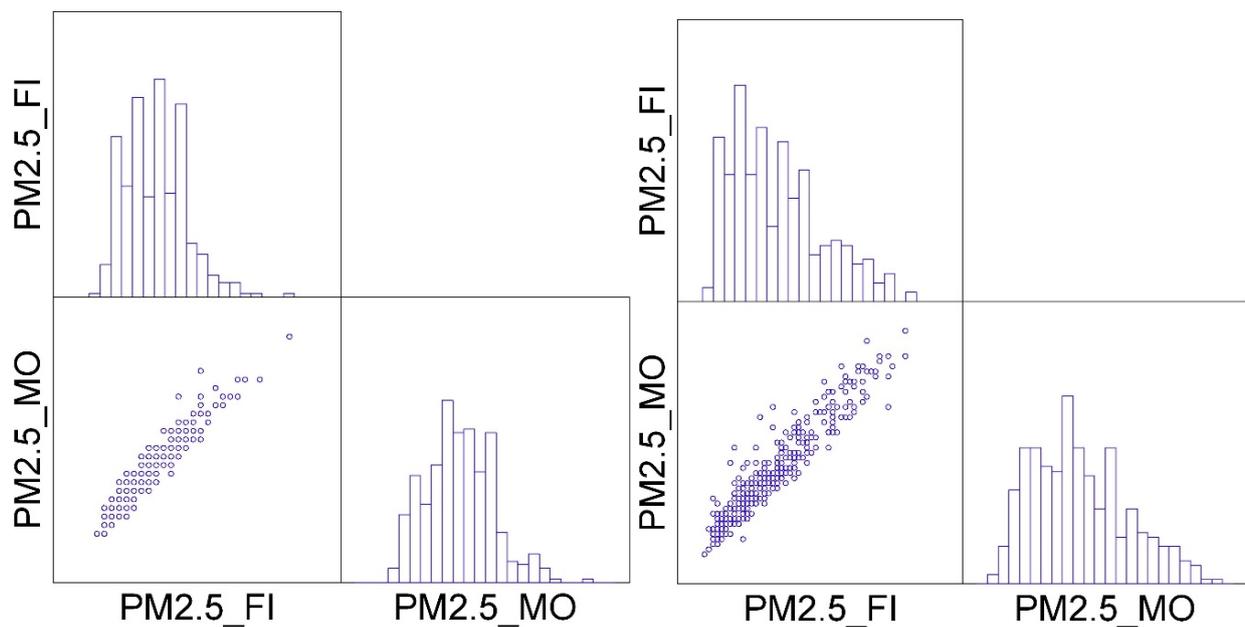**Figure 9.** Scatterplot of PM2.5 measured in the urban background stations TH and CH, summer (**left**) and winter (**right**) semester.

Figure 10. Scatterplot of PM2.5 measured at the traffic stations FI and MO, summer (**left**) and winter (**right**) semester.

Table 4. Summary table of the average of daily values of the PM2.5/PM10 ratio in the eight stations of the network in the two-year period 2017–2018 average value in the summer semester, in the winter semester, and for the whole year.

Period	TH	FI	MO	CH	FR	OR	AQ	CS
April–September	0.50	0.57	0.58	0.62	0.58	0.65	0.62	0.65
October–March	0.70	0.67	0.69	0.74	0.72	0.75	0.69	0.74
year	0.60	0.62	0.64	0.66	0.65	0.70	0.66	0.69

Table 5. Pearson linear correlation coefficient matrix for PM2.5 (summer semester above the diagonal, winter semester below), for the six stations of the coastal strip.

PM2.5	TH	FI	MO	CH	FR	OR
TH		0.934	0.904	0.857	0.922	0.872
FI	0.953		0.954	0.902	0.926	0.902
MO	0.885	0.943		0.882	0.925	0.898
CH	0.923	0.913	0.847		0.874	0.875
FR	0.930	0.948	0.912	0.905		0.891
OR	0.851	0.893	0.847	0.883	0.914	

In order to predict the PM2.5/PM10 ratio, the study continued with model analysis using the FFN network. The FNN was implemented using python with the sklearn (scikit-learn.org) and is schematized in Figure 11: it uses a TANH activation function in two hidden layers and a linear one in the output layer. The FNN was run by varying the number of neurons in the hidden layer (from 1 to 35 neurons) to find the best simulation performance; 30 tests of the model were performed, and the FFN was ran 30 times, during which the weights and bias were varied in turn. To make it reproducible after restarting the machine or in a different machine, we fixed the seed. The input neurons used for model analysis were: (1) carbon monoxide CO, (2) relative humidity (RH), (3) temperature (T), (4) amount of rainfall (RA) (for all stations: FI, MO, OR, TH, and all simulations). The purpose of our analysis was to simulate PM2.5 and PM10 separately in order to better estimate the PM2.5/PM10 ratio. We have used the holdout method to verify the model accuracy on the new dataset (i.e., validation dataset). The data was divided into three series: training (70%), validation (15%), and testing (15%). They were selected using indices initially generated randomly, and then kept fixed for all simulations: in this way, we fixed the selection of the dataset for all simulations, leaving only the weights and the bias variable. The network performance function employed in FFNs (Table 6) [5] is the mean square error (MSE), which controls the optimization of weights and biases during the training process. To select the best simulation among the cases generated with the approach described above, the following statistical parameters have been used and listed in Table 6: the minimum MSE among the target output (i.e., measured PM) and the network output (i.e., modeled PM), the normalized minimum MSE (NMSE) i.e., the MSE divided by the variance of the measured PM, and the maximum correlation coefficient (R) between the measured and modelled PM [5].

Table 6. Scatter plot of modelled vs measured PM2.5/PM10 for the stations (FI, MO, OR, TH) with CO and without CO (no CO) as input neuron. In both cases, the meteorological parameters used as input were the same (T, RH, RA).

	CO					
Station	R	NMSE	FB	FA2	Slope	Intercept
CS	0.45	0.04	0.02	1.03	0.58	0.30
FI	0.63	0.03	0.00	1.01	0.74	0.16
MO	0.57	0.03	0.00	1.01	0.82	0.11
OR	0.55	0.02	−0.01	0.99	0.87	0.08
TH	0.74	0.05	0.02	1.02	0.82	0.11

Table 6. *Cont.*

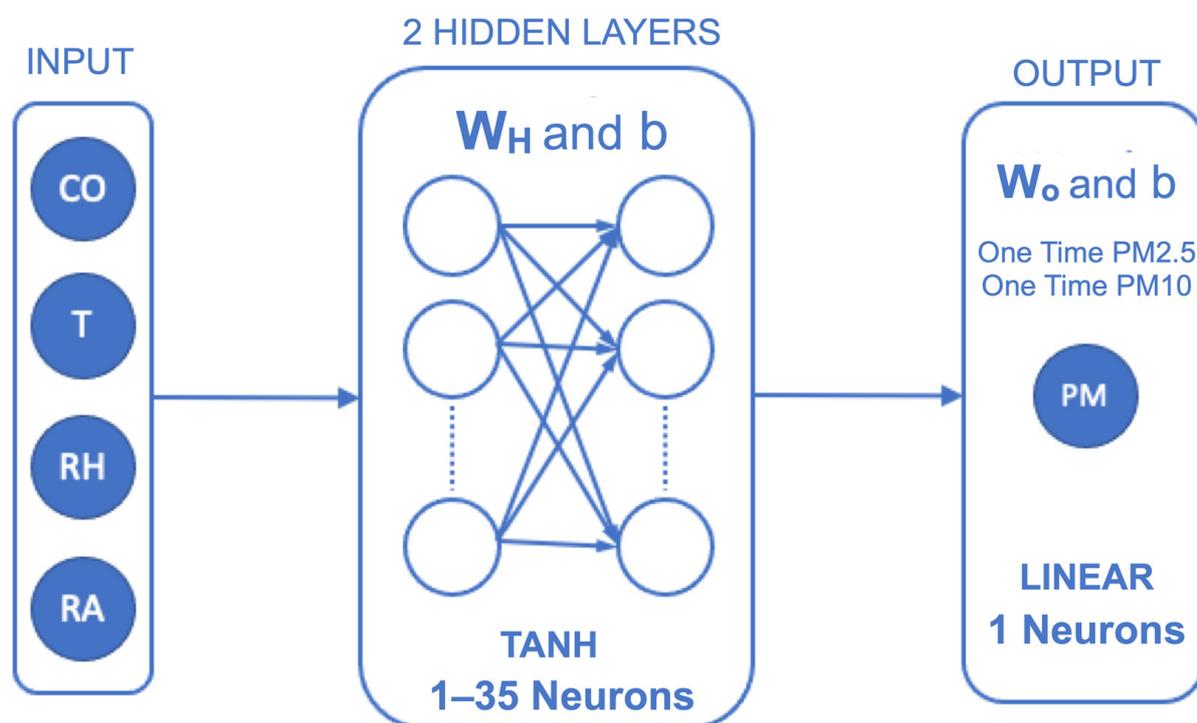|			noCO			
Station	**R**	**NMSE**	**FB**	**FA2**	**Slope**	**Intercept**
CS	0.36	0.04	0.02	1.03	0.60	0.28
FI	0.60	0.03	0.00	1.01	0.75	0.15
MO	0.51	0.03	0.00	1.01	0.65	0.22
OR	0.44	0.03	0.00	1.01	0.45	0.38
TH	0.70	0.06	0.00	1.03	0.81	0.11

Figure 11. Schematic illustration of the feedforward neural networks (FFNs) employed in this study.

2.4. Lin and Pearson Coefficient

An overview analysis of the data has been carried out in order to assess the space-time homogeneity of the PM concentrations measured at the monitoring stations, using the Pearson correlation coefficient [20] together with the Lin concordance coefficient [21,22]. The Pearson correlation coefficient (ρxy) is a measure of the linear correlation between two variables x and y (in this case, daily measurements of the concentration of PM10 or PM2.5 at a pair of stations), which allows to determine whether the two variables (time series of coupled data) are in phase. The Lin coefficient (ρc) is the degree of linear agreement between the two variables, combining precision and accuracy estimates in order to determine how far pairs of experimental data deviate from the line of the perfect agreement [23]. The second stage of the data analysis consisted of the development of machine learning models aimed at predicting the concentration levels of PM10 and PM2.5 as a function of different meteorological variables.

The Lin coefficient, calculated for some of the pairs of stations referred to in Table 5, substantially confirms the results suggested by the Pearson coefficient, highlighting high concordance values only for the FI-MO and TH-FI pairs (Table 7) and, limited to winter for the pair of urban background stations TH-CH, both belonging to the AGG area but located at the two borders of the same (TH is very close to the sea, CH is located in the Pescara valley

15 km from the coast). This analysis confirms that in winter there is a high homogeneity of the spatial distribution of PM2.5 in the AGG. A further analysis was conducted by examining the variation of the Pearson correlation coefficient for PM2.5 measured at the two stations, TH and CH, as a function of the prevailing direction of the wind (without distinguishing between summer and winter). The results show that the highest values (between 0.931 and 0.944) are due to winds from the western quadrants (in particular SSW and SW, the most prevalent directions), while the lower values (between 0.81 and 0.87) occur when the wind blows from the eastern quadrants (NNE to ESE). It should be noted that the SW-SSW winds are directed along the axis of the Pescara valley (in which the main sources of anthropogenic emission of particulate matter are located), from the inside towards the coast, while the winds from the eastern quadrants carry air of marine origin, poor in particulate matter of anthropogenic origin.

Table 7. Lin's coefficient for PM2.5, calculated for some pairs of stations, for both semesters.

PM2.5	FI-MO	MO-FR	FI-TH	FI-FR	TH-FR	TH-MO	TH-CH	CH-OR	TH-OR
April–September	0.951	0.857	0.930	0.862	0.884	0.899	0.851	0.855	0.863
October–March	0.934	0.841	0.947	0.879	0.836	0.861	0.919	0.708	0.685

3. Results and Discussion

In this work we analyzed the data of PM10 and PM2.5, measured in the two-year period 2017–2018, in eight stations of the network, of which five were in the Pescara-Chieti AGG, two were, in the MAXP, and one was in the MINP. As expected, the highest concentrations are found in the AGG. There are also evident seasonal variations, more accentuated for PM2.5, which is more affected by the anthropogenic contribution. The average values of PM10 and PM2.5 are all below the annual limits established by Directive 2008/50/EC of the European Parliament and of the Council of 21 May 2008 (relating to ambient air quality and cleaner air in Europe) (of 40 μg m^{-3} and 25 μg m^{-3}, respectively), and also for the stations of the AGG characterized by a substantial spatial homogeneity of pollution levels.

In this analysis, we used a FFN network to predict the PM2.5/PM10 ratio by using as model input the meteorological parameters that may impact the PM evolution: RH, T, and RA. WS and WD were not used as input parameters in the final simulation because their data showed several gaps in almost all the stations. Therefore, in the model setup, the inclusion of WS and WD significantly reduced the possibility of training and validation of the model. On the other hand, to figure out the role of the ratio of PM2.5/PM10 on the identification of the origin of the PM, FFN network model simulations were carried out in two configurations, including as input the concentrations of CO in the first simulations and excluding CO. In both configurations the meteorological parameters were kept fixed as input. Table 6 and Figure 12 also summarize the results of the model simulations for the station where enough data of CO and meteorological data were available. Comparing the model results for the stations dominated by biogenic emissions (TH and CS), looking at the slope and intercept of observed vs modelled PM2.5/PM10 ratio, the inclusion of CO as input for the model has no effect on the performance of the simulation. Since CO is a good proxy of anthropogenic emission [6] what we found in terms of simulations with and without CO as input means that in sites where biogenic emissions dominate the PM fraction, the PM2.5/PM10 ratio is a good proxy of the PM origin. On the other hand, in sites where the anthropogenic emissions dominate the PM fraction (MO and OR), the inclusion of CO as input in the model results in an improvement of all the statistical parameters such as slope and intercept (Table 6 and Figure 12). These results are further proof that the PM2.5/PM10 ratio can be used as a proxy to classify the PM origin. On the contrary, the FI station that is in an urban area, where we expect anthropogenic emissions to dominate the PM fraction, shows a behavior not consistent with the analysis above, since the inclusion of CO seems to not affect the performance of the model. This result could be explained

considering that FI station is an urban canyon where mixing and evolution of PM, CO, and meteorological parameters are rather complicated and sometimes biased by single local point emission.
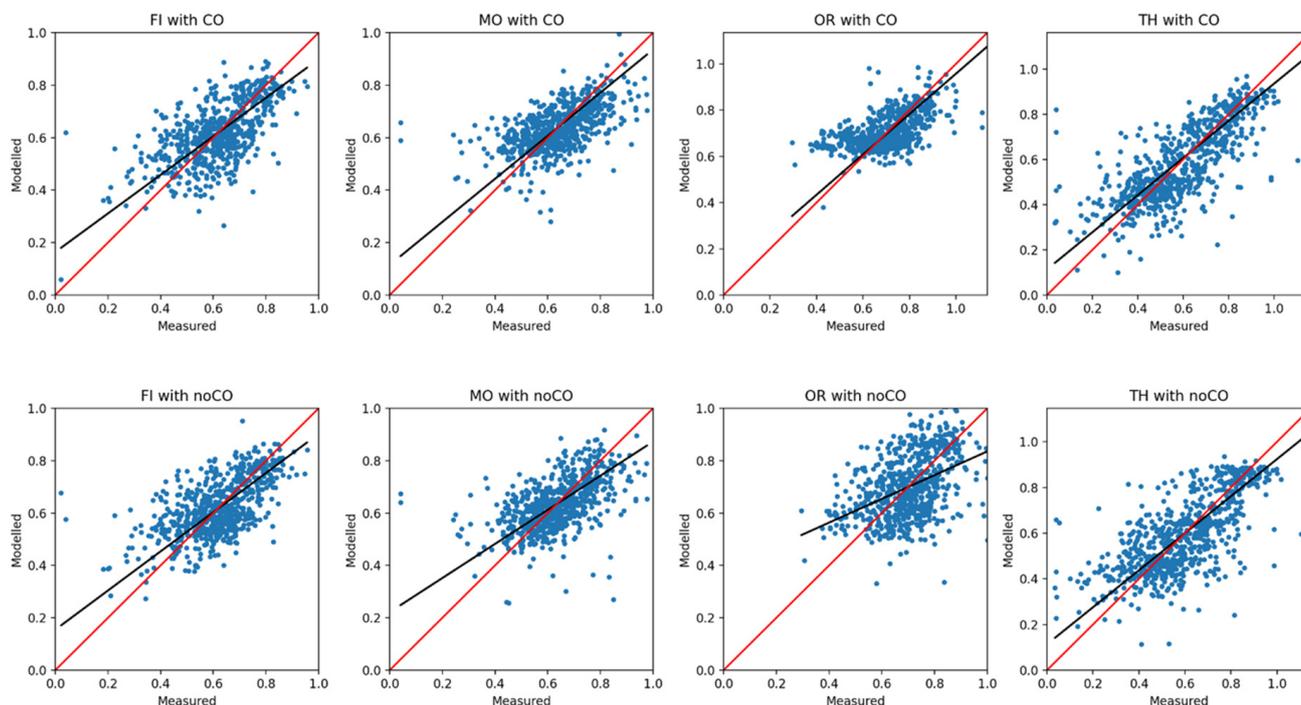

Figure 12. Linear regression model for all stations (CS, CH, FI, FR, MO, OR, TH) without WD and WS.

4. Conclusions

The analysis of the spatial correlation indices between the various stations, within the AGG, confirms a significant spatial homogeneity of the PM concentrations in the winter semester, for which the urban background stations tend to record similar values to those located near the emission sources, for example traffic (the spatial concentration gradients are reduced). On the contrary, the meteorological conditions typical of the summer semester favor the mixing of the boundary layer and the dispersion of pollutants, restoring the spatial concentration gradients and consequently reducing the degree of correlation between stations.

The average values of the PM2.5/M10 ratio provide useful information to understand the possible sources of emissions. In fact, they show marked differences between the various stations in summer, due to the significant contribution of biogenic emission (mainly marine aerosol) affecting the stations located near the coastline. In particular, the TH station, located just 250 m from the coastline, shows a lower PM2.5/PM10 value (R < 5) in June–August compared with other stations (Figure 5). This phenomenon is confirmed both by the trend of the concentrations of PM 2.5 and PM10 in relation to the direction and speed of the wind (as clearly highlighted by the Polar plots, shown in Figure 7a,b) and by the analysis of the model using the network FFN, which showed that the PM2.5/PM10 ratio can be a good tool for analyzing the origin of PM. In detail, the simulations using RA as proxy of the origin of the PM gave, at least for the measurements and the sites analyzed in this work, proof that the PM2.5/PM10 ratio has a completely different behavior in sites dominated by biogenic emissions compared with those dominated by anthropogenic one. This result demonstrates that the analysis of the PM2.5/PM10 ratio, with the help of model analysis using the FFN network, is an excellent tool to know the nature of the emission source (natural and anthropogenic). Further studies are needed to confirm these results, such as simultaneous measurements of the PM2.5/PM10 ratio and the chemical

composition of PM by means of Aerosol Mass Spectrometer (AMS) or Scanning Electronic Microscopy (SEM).

References

1. Liu, Y.; Zhou, Y.; Lu, J. Exploring the relationship between air pollution and meteorological conditions in China under environmental governance. *Nat. Res. Sci. Rep.* **2020**, *10*, 14518. [CrossRef] [PubMed]
2. Kayes, I.; Shahriar, S.A.; Hasan, K.; Akhter, M.; Kabir, M.M.; Sala, M.A. The relationships between meteorological parameters and air pollutants in an urban environment Global. *J. Environ. Sci. Manag.* **2019**, *5*, 265–278.
3. Li, C.; Huang, Y.; Guo, H.; Wu, G.; Wang, Y.; Li, W.; Cui, L. The Concentrations and Removal Effects of PM10 and PM2.5 on a Wetland in Beijing. *Sustainability* **2019**, *11*, 1312. [CrossRef]
4. Arhami, M.; Kamali, N.; Rajabi, M.M. Predicting hourly air pollutant levels using artificial neural networks coupled with uncertainty analysis by Monte Carlo simulations. *Environ. Sci. Pollut.* **2013**, *20*, 4777–4789. [CrossRef] [PubMed]
5. Biancofiore, F.; Verdecchia, M.; Di Carlo, P.; Tomassetti, B.; Aruffo, E.; Busilacchio, M.; Bianco, S.; Di Tommaso, S.; Colangeli, C. Analysis of surface ozone using a recurrent neural network. *Sci. Total Environ.* **2015**, *514*, 379–387. [CrossRef] [PubMed]
6. Biancofiore, F.; Busilacchio, M.; Verdecchia, M.; Tomassetti, B.; Aruffo, E.; Bianco, S.; Di Tommaso, S.; Colangeli, C.; Rosatelli, G.; Di Carlo, P. Recursive neural network model for analysis and forecast of PM10 and PM2.5. *Atmos. Pollut. Res.* **2017**, *8*, 652–659. [CrossRef]
7. Aruffo, E.; Di Carlo, P.; Cristofanelli, P.; Bonasoni, P. Neural Network Model Analysis for Investigation of NO Origin in a High Mountain Site. *Atmosphere* **2020**, *11*, 173. [CrossRef]
8. Chen, J.; De Hoogh, K.; Gulliverd, J.; Ho-Manne, B.; Hertelf, O.; Ketzel, M.; Bauwelinckh, M.; van Donkelaari, A.; Hvidtfeldtj, U.A.; Katsouyannik, K. Comparison of linear regression, regularization, and machine learning algorithms to develop Europe-wide spatial models of fine particles and nitrogen dioxide. *Environ. Int.* **2019**, *130*, 104934. [CrossRef] [PubMed]
9. Cortina-Januchs, M.G.; Quintanilla-Dominguez, J.; Vega-Corona, A.; Andina, D. Development of a model for forecasting of PM10 concentrations in Salamanca, Mexico. *Atmos. Pollut. Res.* **2015**, *10*, 5094.
10. Gardner, M.W.; Dorling, S.R. Neural network modeling and prediction of hourly NOx and NO$_2$ concentrations in urban air in London. *Atmos. Environ.* **1999**, *33*, 709–719. [CrossRef]
11. Mclean Cabanerosa, S.; Calautitb, J.K.; Hughesa, B.R. A review of artificial neural network models for ambient air pollution prediction. *Environ. Model. Softw.* **2019**, *119*, 285–304. [CrossRef]
12. Grimes, D.I.F.; Coppola, E.; Verdecchia, M.; Visconti, G. A neural network approach to real-time rainfall estimation for Africa using satellite data. *J. Hydrometeorol.* **2003**, *4*, 1119–1133. [CrossRef]
13. Shahraiyni, H.T.; Sodoudi, S. Statistical Modeling Approaches for PM10 Prediction in Urban Areas; A Review of 21st-Century Studies. *Atmosphere* **2016**, *2*, 15. [CrossRef]
14. Abdullah, S.; Ismail, M.; Ahmed, A.N.; Abdullah, A.M. Forecasting Particulate Matter Concentration Using Linear and Non-Linear Approaches for Air Quality Decision Support. *Atmosphere* **2019**, *10*, 667. [CrossRef]
15. May Tzuc, O.; Bassam, A.; Ricalde, L.J.; Cruz May, E. Sensitivity Analysis With Artificial Neural Networks for Operation of Photovoltaic Systems. In *Artificial Neural Networks for Engineering Applications*; Alanis, A.Y., Arana-Daniel, N., López-Franco, C., Eds.; Academic Press: Cambridge, MA, USA, 2019; Volume 10, pp. 127–138.
16. Sairamya, N.J.; Susmitha, L.; Thomas George, S.; Subathra, M.S.P. Hybrid Approach for Classification of Electroencephalographic Signals Using Time-Frequency Images With Wavelets and Texture Features. In *Intelligent Data Analysis for Biomedical Applications*; Hemanth, D.J., Gupta, D., Balas, V.E., Eds.; Academic Press: Cambridge, MA, USA, 2019; Volume 12, pp. 253–273.

17. Castro, W.; Oblitas, J.; Santa-Cruz, R.; Avila-George, H. Multilayer perceptron architecture optimization using parallel computing techniques. *PLoS ONE* **2017**, *12*, 0189369. [CrossRef] [PubMed]

18. Guo, Z.; Chai, Q.; Maskell, D.L. FCMAC-AARS: A Novel FNN Architecture for Stock Market Prediction and Trading. In Proceedings of the IEEE International Conference on Evolutionary Computation, Vancouver, BC, Canada, 16–21 July 2006; pp. 2375–2381.

19. Cattani, G.; Di Menno Di Bucchianico, A.; Gaeta, A.; Gandolfo, G.; Leone, G. Qualità dell'ambiente urbano. XIV Rapporto ISPRA Stato dell'Ambiente 82/18. *Riv. Ital. di Econ. Demogr. e Stat.* **2018**, *5*, 375–441. Available online: https://www. isprambiente.gov.it/it/pubblicazioni/stato-dellambiente/xiv-rapporto-qualita-dell2019ambiente-urbano-edizione-2018 (accessed on 20 January 2022).

20. Zauli Sajani, S.; Scotto, F.; Lauriola, P.; Galassi, F.; Montanari, A. Urban Air Pollution Monitoring and Correlation Properties between Fixed-Site Stations. *J. Air Waste Manag. Assoc.* **1994**, *54*, 1236–1241. [CrossRef] [PubMed]

21. Biggeri, A.; Baccini, M.; Accetta, G.; Bellini, A.; Grechi, D. Valutazione di qualità delle misure di concentrazione degli inquinanti atmosferici nello studio dell'effetto a breve termine dell'inquinamento sulla salute. *Epidemiol. Prev.* **2003**, *27*, 365–375. [PubMed]

22. Palermi, S.; Polidoro, M.; Di Tommaso, S.; Colangeli, C.; Bianco, S. Omogeneità spaziale delle concentrazioni di Benzo(a)Pirene misurate presso due stazioni nell'area urbana di Pescara. *Boll. Degli Esperti Ambient.* **2016**, *3*, 45–58.

23. Lawrence, I.; Lin, K. A Concordance Correlation Coefficient to Evaluate Reproducibility. *Biometrics* **1989**, *45*, 255–268.