Politecnico Di Milano

Environmental and Geomatic Engineering

"Duration of the PM2.5 particles in Lombardy and Macedonia"

Supervisor: prof. Giovanni Lonati
Supervisor: prof. Elena Sezenna

Master Graduation Thesis by:

Stefan Velkovski Id. Number 836459
Master of Science in Environmental and Geomatic Engineering
POLO REGIONALE DI COMO

Academic Year 2016/2017

Table of Contents

List of figures

List of tables

Abstract

For this case-study annual time series consisting of daily average concentrations of fine particulates known as PM2.5 are used. The yearly concentrations are consisted into separate data sets. Measuring was done in 11 stations across Lombardy and Macedonia, for a given time-period of 5 years, from 2012-2017. Analysis performed include duration of events where the concentration of pollutant is exceeding certain threshold, where the event is represented by a series of days when the fixed threshold is exceeded without interruption. In order to calculate the total time (TE_T), number of events (NE_T) and the duration of the events of exceedance annually, 10 to 16 concentration thresholds ranging from 10 μg m^{-3} to 90 and 150 μg m^{-3} with a step of 10 μg m^{-3}, and a special threshold of 25 μg m^{-3} were used. Followed by parameterized empirical or theoretical models of the yearly data, made to explain the observed data (NE_T, TE_T, and the duration of the events of exceedance), and can be further used to estimate concentration levels in future scenarios. Due to the higher concentration levels of PM2.5 than the one regarded as limit by the EU, this case-study including Lombardy and Macedonia makes assessments of the frequency of the duration episodes in order to attain a target average concentration limit for PM2.5 where the annual concentration will be in compliance with the limit.

1 Introduction

Air pollution has posed problems since the beginning of the age of burning coal. Through the industrial revolution in the 18th and 19th century, and the mass continuous industrialization in the 20th century. This has increased the GHG in the atmosphere and through the combustion the concentration of the secondary pollutants such as the PM have reached concerningly high levels. Not until the 1970s the scientists have begun to search for a relationship between the air pollution and the cardiovascular diseases (EPA, 2017).

The ongoing problems caused by the pollution have not yet subsided. Scientist have begun to enforce policies and legislations for the abatement of air pollution, through implementation of Air Quality Standards, that show the quality of the air presented as one number. These legislatives provide a way of action once a concentration limit (threshold) of a given pollutant has been exceeded for many consecutive days. In the developed countries in the last years the pollution has almost subsided but the success of the legislatives is yet to be seen in the developing countries.

Pollution levels in EU are within the annual limit but still there is room for improvement. On the other hand, the countries in development have not yet reached the minimum concentration limit of the pollutants that is required. In order for these countries to do so they need to figure out what is the cause of high emissions of air pollution and to reduce it.

1.1 Particulate Matter

What is known as Particulate Matter (PM) or just particulates are the sum of all extremely small solid and liquid matter suspended in the air of the Earth's atmosphere in comparison to the term aerosol which represents the mixture of air/particulate. (Seinfeld & Pandis, 1998) It includes a variety of chemical combination of compounds like metals, organic components, dust/soil particles, sulfates/nitrates. (Kelly, 2012) Examples include dust, ash and sea-spray. PM is discharged during combustion of liquid and solid fuels. Varies in size (i.e. the diameter of the particle). The PM are divided into two groups. PM2.5 are particles with diameter smaller than 2.5 µm or fine particulates which signifies the mass per cubic meter of air of particles having diameter (width) less than 2.5 micrometers (µm), and PM10 particles with diameter bigger than 2.5 µm and smaller than 10 µm, see Fig. 1. Furthermore, the particles with size ranging from 0.1-50 µm are called TSP or total suspended particles. The smaller the size of the particles the easier they are to be infiltrated into the farthest parts of the lungs and even into the bloodstream.

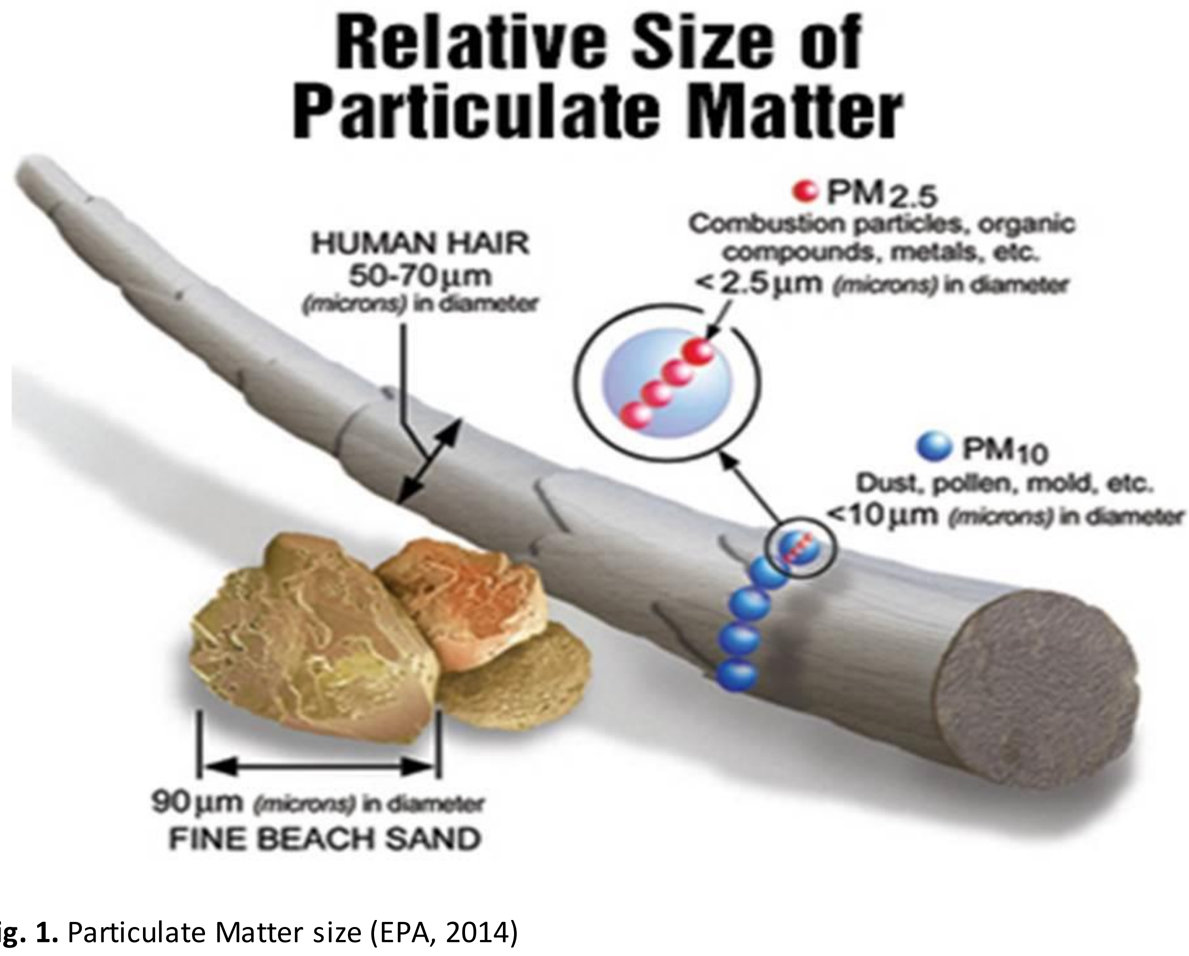


Fig. 1. Particulate Matter size (EPA, 2014)

PM10 particles are coarse and they are made usually by combustion and through breaking of bigger particles. These particles consist of different materials depending on their location. PM10 contains pollen and bacteria, dust from roads and industrial areas and soil.

PM2.5 or fine particles are released in the atmosphere naturally or anthropogenically. The human-made sources' contribution to the total concentration is very significant in contrast to the natural sources. They are produced from a wide range of industrial processes through bulk material handling, combustion and minerals processing, however they can also be produced from gases.

The sources of PM2.5 can be primary and secondary.

Primary sources directly emit the pollutant into the atmosphere. They are as follows:

- Motor vehicles
- Industrial processes (power plants)
- Natural sources (forest fires, bushfires, dust storms, pollens and sea spray)
- Local sources
- Residential wood burning
- Circulation of pollution on a given region, known as Long range transportation of air pollution
- Long range transport + local source pollution = short term episodes of high pollution
- All types of combustion

Motor vehicles emit high concentration of pollutant in the local area around the roads. The residents living near traffic areas will experience higher pollution than those living in rural areas.

Industrial regions' emission of PM2.5 is more significant than the road vehicles in these regions depending on the fuel used for combustion and the meteorological conditions. The polluted air can be circulated to another region due to meteorological conditions and the phenomenon of long range transportation of air pollution can be experienced.

The short-term episodes of high pollution can be very hazardous to the human health especially to those that are sensitive to it, if the concentration of the pollutant is extremely high.

Secondary sources are formed in the atmosphere as a result of a chemical reaction between gaseous pollutants. They are:

- Sulphur dioxide (SO_2)
- and nitrogen oxides (NOx: nitric oxide, NO plus nitrogen dioxide, NO_2)

These are known as secondary particles.

In order to reduce the levels of PM2.5 one must reduce the emissions of the precursor gases.

In one country, the total concentration of PM2.5 is the sum the of the primary emissions of PM2.5, the formation of secondary PM within the country and the long transport of pollution from outside the country.

As the particles get smaller in size, their time passed in the air increases. From few minutes or hours to few days or weeks, for PM10 and PM2.5 respectively. Another thing is the travel distance, between the particles, again the smaller particles could travel up to hundreds of kilometers while the bigger ones could only go to few hundred meters (Stephanou, 2012).

Health effects from overexposure of particulates can be acute (short-term) or chronic (long-term). Acute health effects begin to appear after quick exposures to very high levels or concentrations of a pollutant, in this case particulate matter. Whereas chronic health effect begins to show after long term or repeated exposure to a certain pollutant for over a given time-period (day, week, month, year). Heath effects could be reversible if one is exposed to a single event of high levels of a pollutant, but irreversible when one is exposed to high levels of pollutant for a long period of time. (OSHA, 2017).

As seen from the Fig. 2, the ambient particulate matter pollution is in the first 5 leading causes of death in the world.

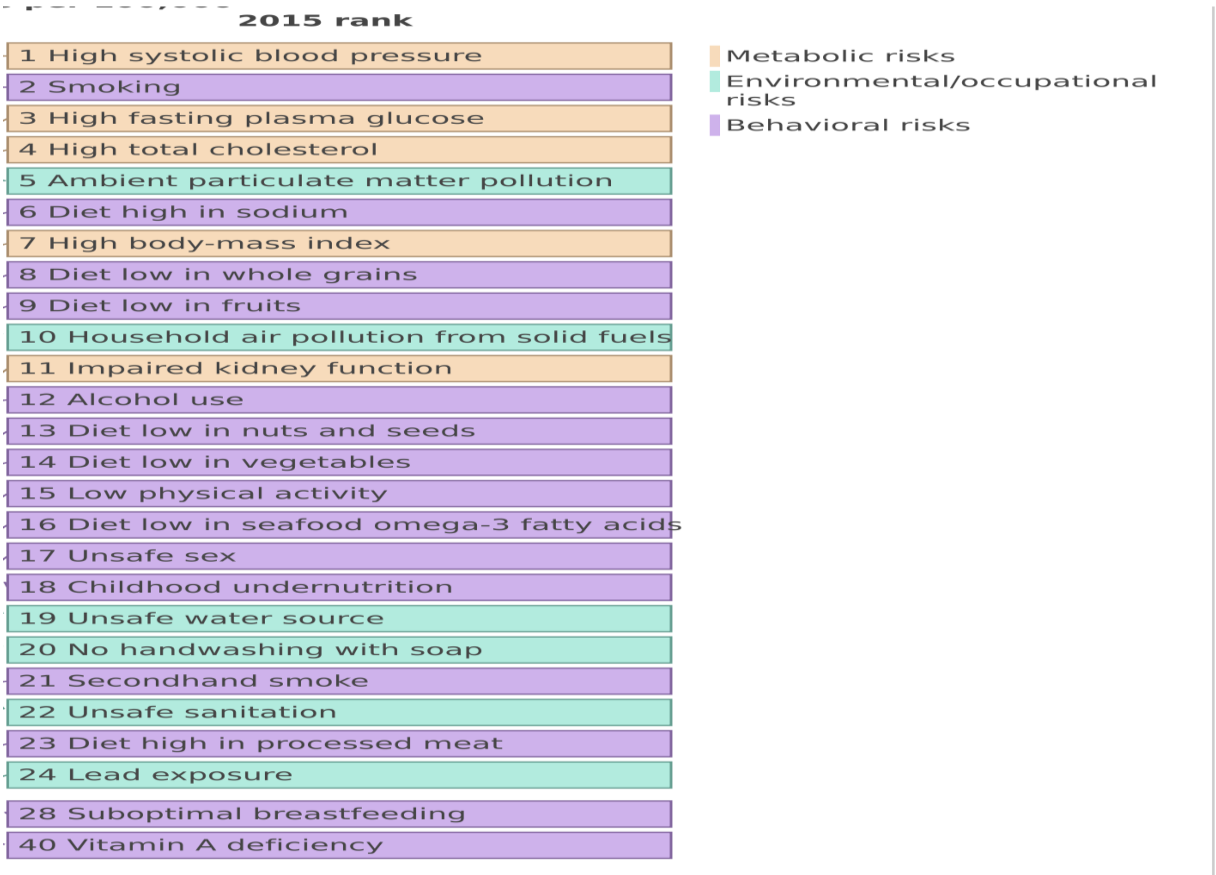


Fig. 2. Global Causes of Death (Lancet, 2016))

There is a link between cardiovascular diseases and the PM pollution as many epidemiological studies suggest. (Lancet, 2016) What was found is that long-tern exposure to fine PM could increase the morbidity and mortality and so reducing the life expectancy. (Pope CA 3rd, 2009)

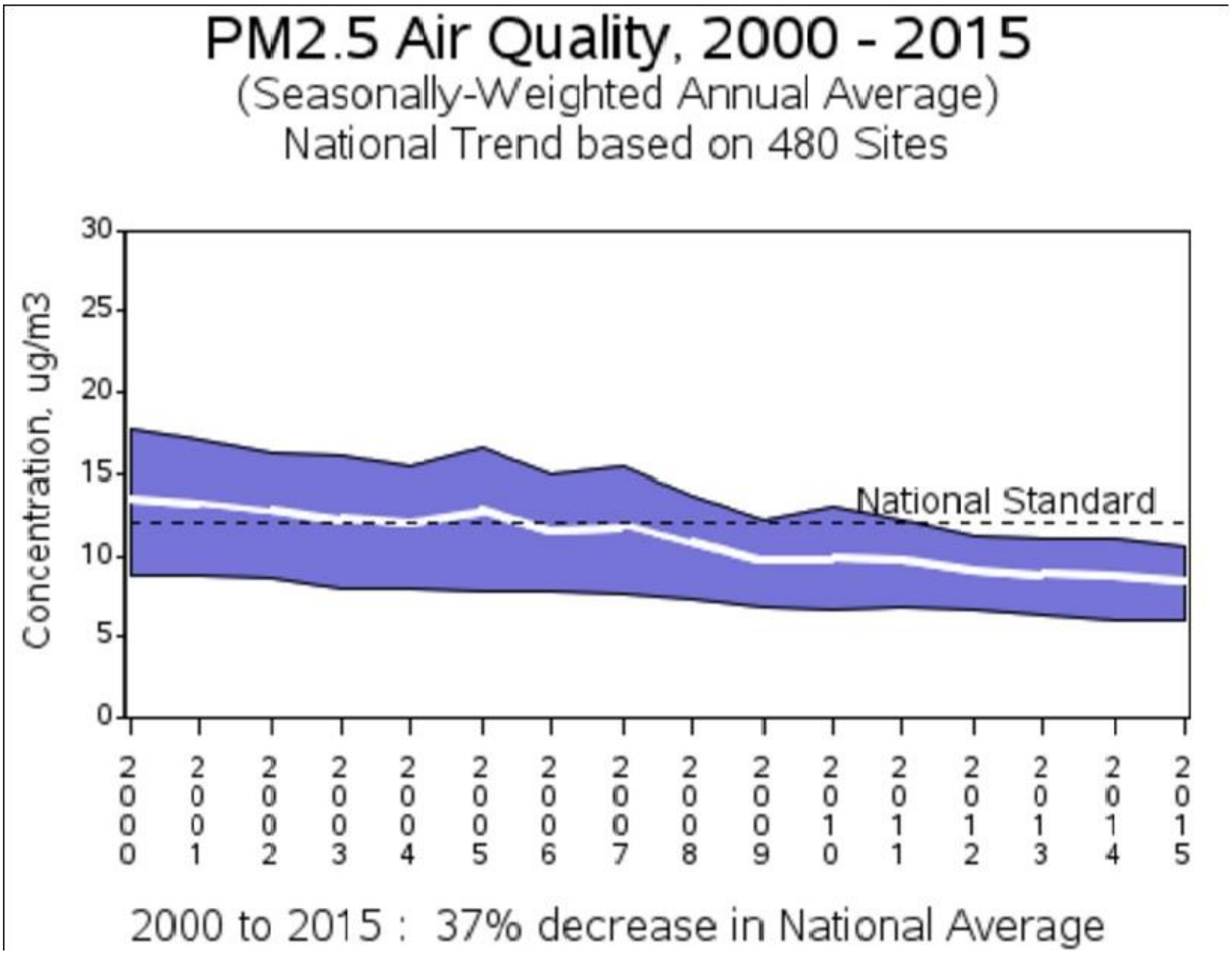


Fig. 3. PM2.5 concentration from the year 2000 to 2015

Even though the concentration of PM2.5 is being reduced over the years (falling trend, see Fig. 3), there is still a lot of under developed countries that do not follow the policies for PM abatement. The scientists are still trying to find the minimum concentration level of PM2.5 below which the pollution will not cause any adverse effects on the human health. But there is another thing that needs to be considered, that not all human beings react the same. What this signifies is the following, depending on the age, gender, genetics and life choices, the individuals will react differently to the same level of pollution, due to the complexness of the human organism. For example, the people living in poor socioeconomic classes may suffer more since they are already facing poor conditions. At the end, the effect of pollution could be correlated with other influential elements (S. Koton, 2013).

PM10 and PM2.5 particles are present in the air. From there they can be easily breathed in through the nose/mouth into the lungs. Some of the PM may enter even in the bloodstream, depending on the size, the smaller PM2.5 may reach the farthest parts of the lungs, and get stuck there. As known they carry toxic compounds and are far more hazardous to the human health than the bigger PM10. The problems caused by the PM2.5 are the following:

- inflammation, redness of eyes, nose and throat
- arrhythmia
- reduced lung capacity and function, coughing

- asthma attacks
- rare case of death in people with lung disease

(S. Koton, 2013).

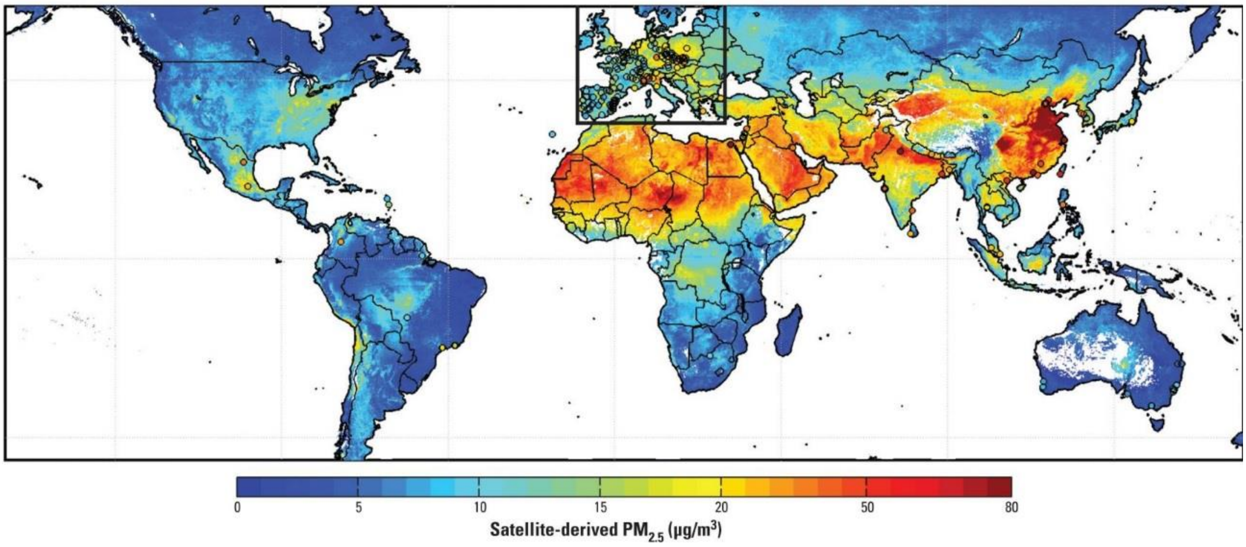


Fig. 4. Global satellite-derived PM2.5 averaged over 2001–2006. White space indicates water or locations containing < 50 measurements.

As mentioned before PM2.5 causes damage to the humans as well as the nature. On the Fig. 4 one can see the global PM2.5 averaged between 2001-2006.
Environmental damages caused by the fine particles are as follows:

- acid rain
- haze

Acid rain is a very dangerous phenomenon caused by natural or anthropogenic sources. The water in the air reacts with the Sulphur dioxide and nitrogen oxide creating nitric and sulfide acids, which when mixed with the water in the atmosphere becomes acid rain. This causes harmful effects on all living beings, such as plants, and on any other material, such as rocks. Acidic rain causes stripping of nutrients in the soil thus destroying any chance of vegetation on that location. Due to the small size of the PM2.5 the wind could blow them away to another location, thus relocating the acid rain.
Haze is an atmospheric phenomenon where smoke, dust or other particles reduce the visibility. (Sun, 2014))
In the end, to reduce the effect of any, and all sources of pollution one needs to know the exact source that causes the problem (ex. Domestic fires). Once that is done one needs to notify the people about the source and to tell them how to reduce the emissions (ex. usage of less wood for burning).

1.2 Air quality limits for PM

As said before the PM represents the sum of PM10 and PM2.5. Because of the difference in the diameter of the particles different AQL exist for PM10 and PM2.5.
The air quality limits for PM are according to the European directive (Union, 2008) equal to 50 µg m^{-3} but cannot be exceeded for more than 35 times throughout the year, seen on Fig. 5.

PM$_{10}$			
One day	50 µg/m^3, not to be exceeded more than 35 times a calendar year	50 %	— (2)
Calendar year	40 µg/m^3	20 %	— (2)

Fig. 5. PM10 air quality limit (Union, 2008)

For the PM2.5 the air quality limit is 2 times lower, equivalent to 25 µg m^{-3} because of the smaller diameter. For the PM2.5 there is no limit for exceedances throughout the year.

For the following years the European directive (Union, 2008) has made suggestions for further reduction of the PM2.5 from 25 µg m^{-3} to 20 µg m^{-3} by the start of 2020, seen on Fig.6.

E. Limit value

Averaging period	Limit value	Margin of tolerance	Date by which limit value is to be met
STAGE 1			
Calendar year	25 µg/m^3	20 % on 11 June 2008, decreasing on the next 1 January and every 12 months thereafter by equal annual percentages to reach 0 % by 1 January 2015	1 January 2015
STAGE 2 (1)			
Calendar year	20 µg/m^3		1 January 2020

Fig. 6. PM2.5 air quality limit (Union, 2008)

Furthermore, the World Organization of Health (WHO, 2005) suggests an even smaller limit of 10 µg m^{-3} for which it is concluded that the adverse effects caused by the PM2.5 are reduced to minimum. However, even if a country meets this limit adverse effects on the human health are still a possibility because of the background levels and natural sources of these pollutants.

1.3 Emission inventory

For this case-study two different regions will be compared, Macedonia and Lombardy. Macedonia is a country in development with an area of 25 713 km^2, and population of 2.1 million. Whereas Lombardy is spread on almost the same area 23 844 km^2, but the population living is 5 times greater, reaching a number of 10 million inhabitants. When comparing the emissions of PM one expects to see lower numbers for Macedonia in contrast to Lombardy. Here that is not the case. Emissions in both regions are almost the same.

The Table 1 bellow shows how much each sector emits PM2.5 and PM10 in Lombardy and Table 2 shows the TSP emitted in Macedonia. The TSP represents the Total Suspended Particles, including both PM2.5 and PM10. From the EU – Emission report 1990-2011 (Union, 2013), page 54, one could work out the ratio between the PM and the TSP. It is as follows:

- PM10/TSP = 0.55
- PM2.5/TSP = 0.36

According to these formulations one could make comparisons between Lombardy and Macedonia, from the Tables 1 and 2. The TSP emission in Macedonia ranges from 15000 to 43000 tonnes per year. From those values, by using the formulations above stating that the composition of PM10 in TSP is 55% and that the composition of PM2.5 in TSP is 36%, one could calculate the exact value of the particulates.

For the minimum and maximum values of TSP in Macedonia one gets the following PM10:

15 996 tonnes/year TSP * 0.55PM10/TSP = 8797.8 tonnes/year PM10

43 351 tonnes/year TSP * 0.55PM10/TSP = 23898.05 tonnes/year PM10

For the minimum and maximum values of TSP in Macedonia one gets the following PM2.5:

15 996 tonnes/year TSP * 0.36PM2.5/TSP = 5758.6 tonnes/year PM2.5

43 351 tonnes/year TSP * 0.36PM2.5/TSP = 15606.4 tonnes/year PM2.5

When compared to the values of PM10 and PM2.5 from Lombardy it could be concluded that the highest TSP emission in Macedonia results in lower values for PM10 and PM2.5 as the ones in Lombardy.

Table 1 Emission inventory Lombardy (Caserini, 2004)

Group	Sub-group	PM2.5	PM10
2 Non-industrial combustion plants	2 Residential plants	4,011	4,361
7 Road transport	3 Heavy duty vehicles > 3.5 t and buses	3,749	4,111
7 Road transport	1 Passenger cars	3,098	3,787
10 Agriculture	10 Animal husbandry	738	3,653
4 Production processes	2 Processes in iron and steel industries and collieries	883	1,784
7 Road transport	2 Light duty vehicles < 3.5 t	1,209	1,278
10 Agriculture	3 On-field burning of stubble, straw,...	1,010	1,182
3 Combustion in manufacturing industry	1 Comb. in boilers, gas turbines and stationary engines	1,006	1,084
8 Other mobile sources and machinery	6 Agriculture	994	1,046
1 Combustion in energy and transformation ind.	1 Public power	656	920
3 Combustion in manufacturing industry	3 Processes with contact	450	807
1 Combustion in energy and transformation ind.	3 Petroleum refining plants	592	658
11 Other sources and sinks	3 Forest and other vegetation fires	511	549
4 Production processes	6 Processes in wood, paper pulp, food, drink and other ind.	118	491
11 Other sources and sinks	25 Other (tabacco smoking and fireworks)	459	459
6 Solvent and other product use	3 Chemical products manufacturing or processing	159	437
2 Non-industrial combustion plants	1 Commercial and institutional plants	252	278
7 Road transport	4 Mopeds and Motorcycles < 50 cm3	111	111
9 Waste treatment and disposal	2 Waste incineration	56	57
8 Other mobile sources and machinery	3 Inland waterways	38	41
7 Road transport	5 Motorcycles > 50 cm3	34	39
4 Production processes	5 Proc. in organic chemical industr. (bulk production)	34	39
4 Production processes	1 Processes in petroleum industries	11	26
8 Other mobile sources and machinery	8 Industry	24	24
8 Other mobile sources and machinery	5 Air traffic	19	20
8 Other mobile sources and machinery	2 Railways	19	20
6 Solvent and other product use	1 Paint application	20	20
4 Production processes	3 Processes in non-ferrous metal industries	10	11
4 Production processes	4 Processes in inorganic chemical industries	7	10
2 Non-industrial combustion plants	3 Plants in agriculture, forestry and aquaculture	10	10
Others		9	16
Total		**20,299**	**27,331**

Table 2 Emission Inventory Macedonia (MOEPP, 2015)

8.6 Total emission of TSP by sector

тони/година tonnes/year

	Вкупно Total	Согорувачки процеси Combustion processes	Производствени процеси Production processes	Транспорт Transport	Останато Other
2006	38 895	3 079	33 148	2 668	-
2007	29 921	6 208	23 487	225	1
2008	23 862	6 007	17 826	20	9
2009	27 392	6 283	20 913	20	176
2010	15 996	9 908	5 425	658	5
2011	23 428	17 316	5 452	658	5
2012	34 893	15 936	18 425	514	18
2013	43 451	16 385	25 713	572	780

Извор: Министерство за животна средина и просторно планирање
Source: Ministry of Environment and Physical Planning

1.4 Preview of the work done

This case-study examines the time series of PM2.5 data from Lombardy and Macedonia.
Up until recent years only PM10 case-studies were done, because data for PM2.5 was either not measured or not required by the EU to be studied. Now EU regulations exist not only for PM10 but also for PM2.5 and so the focus now falls on the PM2.5 data.

The stations from which the PM2.5 data was obtained were situated for the most part in Lombardy and Macedonia. Data sets range from 2012 to 2017. The name and the average concentration can be seen on the Table 3.

What the time series show is the fact that the pollution done by the PM2.5 exceeds the annual limit. 10 to 16 arbitrary concentration thresholds with a step of 10 µg m^{-3}, ranging from 10 – 90 -150 µg m^{-3} were appointed. Showing us the total time of exceedance, how many times the limit has been exceeded and for how long.

In the end based on these calculations parameterized model describing the PM2.5 concentrations has been done, that allows us to predict the duration of PM2.5 concentrations for future scenarios.

2 Description of data series

Data used for this case-study are the daily average concentrations of PM2.5. The time-series were downloaded from two sites, one for Lombardy data (ARPA, 2017), and for Macedonian data (MOEPP, 2017). For reaching a greener future there needs to be a legislative like the EU regulation stating that the limit of the PM2.5 annual average concentration should not exceed 25 µg m^{-3} (Union, 2008), there is also a Macedonian legislation, however it is written only in Macedonian stating the exact limit as in the EU (MOEPP, 2014). Because of the harmful effects of the pollutants released in the atmosphere, people are getting concerned about their health. The exposure to the pollutants and its duration plays the biggest role in determining the adverse effects caused by those pollutants (Ryan, 1991). Every physical phenomenon could be measured and/or modeled, the same thing goes for the pollution. However, when it comes to modelling the atmospheric pollution one needs to account for many things: how long are the people subject to the pollution i.e. how long is the duration of exposure, what is the concentration of the pollutants, how the pollution behaves as a function of time (Ryan, 1986). One could use a combination of Statistical role back modeling and probability model to predict the time of exceedance and the return period of the highest concentrations lacking the time pattern of those concentration levels (G. Lonati, 2011).

To predict the duration of a certain pollutant in the atmosphere is very difficult simply because the concentrations of that pollutant are affected by the meteorological conditions, and the modeling should account for them. When one has, data obtained from measuring the air at monitoring stations, then an analysis could be made. It could be done by fixing concentration thresholds, and characterizing the data by how much they continuously exceed those thresholds.

If one fixes a certain concentration threshold, then one could compute the following elements:

- Total time of exceedance (TE_T)

TE_T represents the total time, in this case-study days, because the gathered data is average daily concentrations of PM2.5, when a fixed concentration threshold is exceeded.

- Number of events of exceedance (NE_T)

NE_T represents the number of events of exceedance, a dimensionless variable, describing how many events exceed a fixed concentration threshold, where an event is represented by a series of days when the fixed threshold is exceeded without interruption.

- Duration of events of exceedance

Duration of events of exceedance shows how long the events lasted, in days.

In order to calculate the total time(TE_T), number of events(NE_T) and the duration of the events of exceedance annually, 10 to 16 concentration thresholds ranging from 10 µg m^{-3} to 90 and 150 µg m^{-3} with a step of 10 µg m^{-3}, and a special threshold of 25 µg m^{-3} were used. For the TE_T and NE_T for each threshold only one value is obtained, but for the duration of the events more than one value for each threshold is obtained. This is due to the fact that this element explains the duration of each event. NE_T and TE_T can be easily represented, whereas for the duration of the events one needs to use distributions to represent all the values for each concentration threshold. These 3 elements will be used in making parameterized empirical models of the annual observations in order to be able to make predictions of the concentrations for future scenarios.

For this case-study data for Macedonia and Lombardy concerning PM2.5 average daily concentrations were used. They ranged from 2012-2017, the number of monitoring stations is 11, so the total number of yearly time-series is 55. The 3 elements previously mentioned (TE_T, NE_T, and the duration of events of exceedance) were analyzed and then computed, so as to determine how many times and for how long there is exceedance of a fixed threshold for annual average PM2.5 data. The location of the monitoring stations in Lombardy and Macedonia can be seen on Fig. 7 and Fig. 8, respectively.

Fig. 7. Lombardy monitoring stations

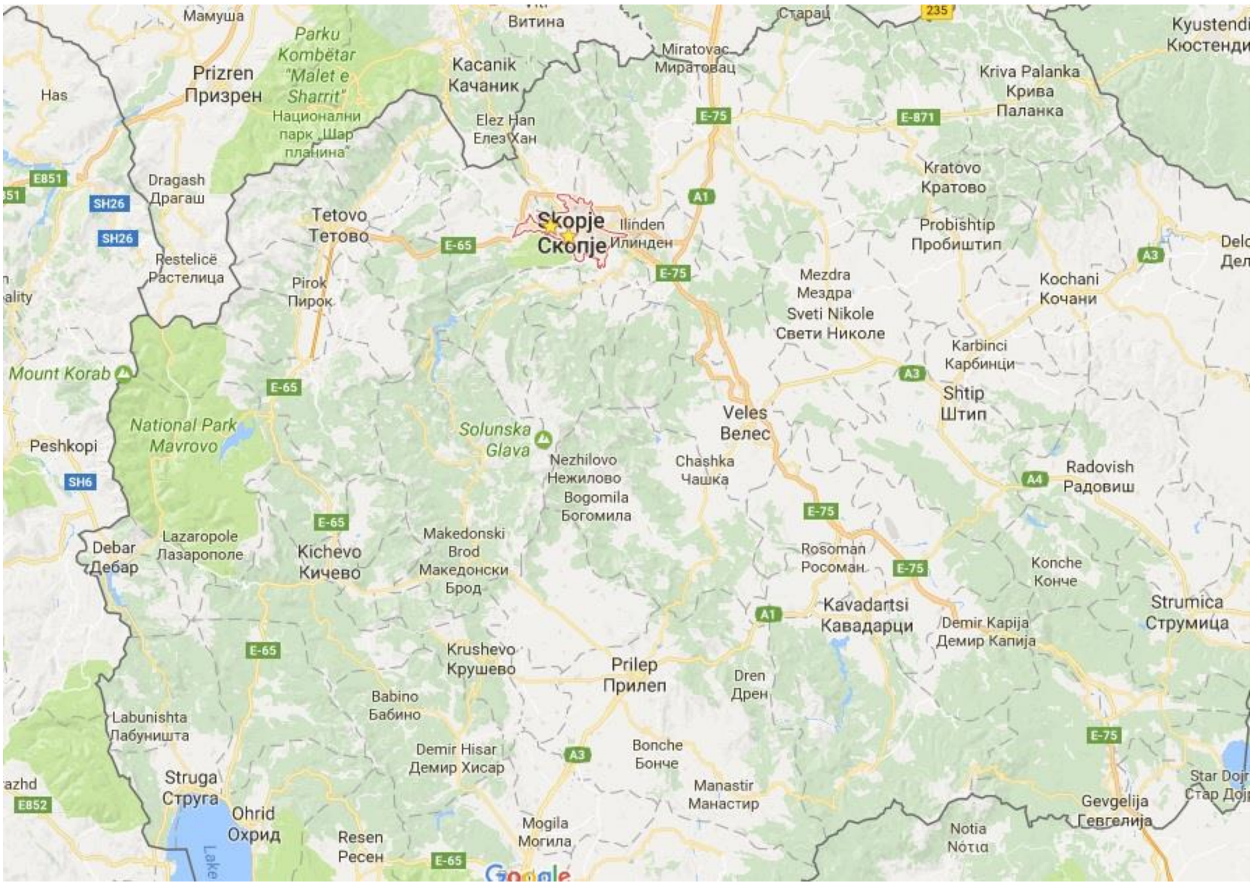


Fig. 8. Macedonia monitoring stations

2.1 Average Concentrations of PM2.5

After downloading and organizing the PM2.5 data from Lombardy and Macedonia one can see on Table 3 the annual data series range in the middle row, and the average annual concentration range in the last row. For this case-study a total of 55 annual data sets were recorded between 2012 and 2017 at 11 stations. Data consistency was a big deal in choosing the stations. The days when there was no average concentration of PM2.5 available (missing data), interruption of the time series was recorded. The missing data was usually due to sensor malfunction. For the case-study only the data sets having below 10% missing data were considered. However, some of the listed stations started monitoring the PM2.5 concentration mid-year, but still the data consistency is not below 50%, and are taken into consideration.

Upon inspection, the two big cities of Skopje and Milan are in fact the most polluted ones. The annual concentration ranges in Lombardy are usually around the acceptable required limit of 25μg m^{-3}, but that is not the case for the two stations in Macedonia Centar and Karpos, they are the ones with the highest annual average concentration equaling to 50 μg m^{-3} which is double than the acceptable limit. Right behind is a station in Milan, Milano Pascal having the highest annual average concentration of 31.94 μg m^{-3}. In this case-study a comparison of few parameters between the most polluted areas of each region will be examined, as well as a development of a model for future scenarios built by data from Lombardy and Macedonia. Those areas as can be seen are Centar and Milano Pascal.

Skopje has a rough estimate of half a million inhabitants whereas Milan has 1.3 million inhabitants. Again, due the underdevelopment of the city of Skopje the pollution is 2 times greater than a city with 2.5 times more people.

Table 3 Details on monitoring stations and PM2.5 data series, the first nine stations are from Lombardy and the last two are from Macedonia

Station location	Annual data series	Average annual concentration range (μg m^{-3})
Bergamo C'Alusco	2012 – 2017	17.6 - 24.6
Bergamo Casirate d'Adda	2012 – 2017	19.8 - 28.6
Bergamo Dalmine	2012 – 2017	24.1 - 28.9
Bergamo Meucci	2012 – 2017	20.0 - 26.9
Bergamo Seriate	2012 - 2017	20.0 - 26.5
Como Cattaneo	2012 – 2017	18.3 - 26.3
Milano Castano Primo	2012 – 2017	20.0 - 26.1
Milano Pascal	2012 – 2017	25.8 - 31.9
Milano Senato	2012 – 2017	22.4 - 29.8
Karpos	2012 – 2017	23.8 - 51.6
Centar	2012 – 2017	25.1 - 52.4

On the Fig. 9. annual concentration of PM2.5 per country based on the monitoring stations can be seen. Where it is obvious that Macedonia has a pollution problem because the annual concentrations of PM2.5 are well beyond the annual average limit proposed by the EU of 25 μg m^{-3}.

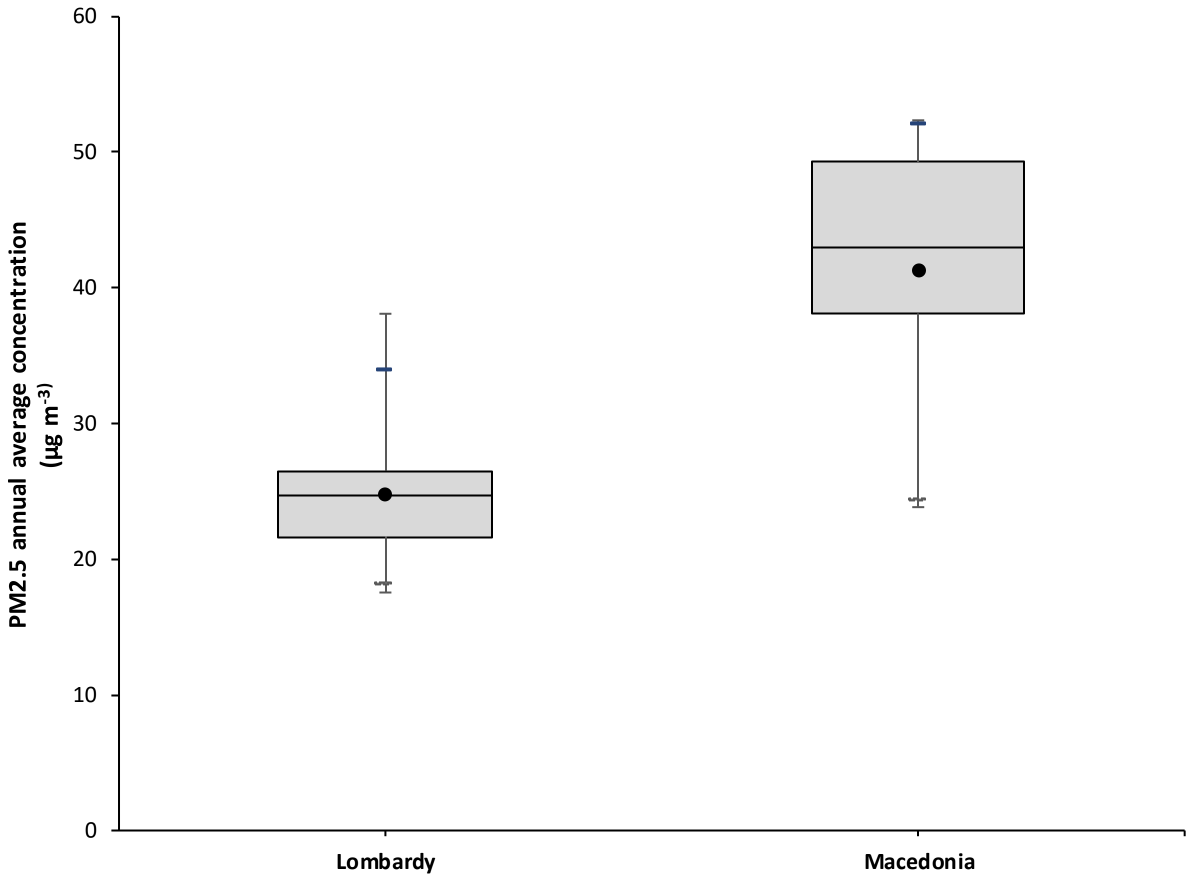


Fig. 9. PM2.5 annual average concentration (μg m^{-3}) at monitoring sites in the two regions (average: black dot; IQR: gray box; 5th & 95th percentile: dash, minimum and maximum value: extreme values).

2.2 Calculation of the total time of exceedance, and number of events

Analysis performed include duration of events where the concentration of a pollutant is exceeding certain concentration threshold. In order to calculate the total time (TE_T), number of events (NE_T) and the duration of the events of exceedance annually, 10 to 16 thresholds ranging from 10 µg m^{-3} to 90 and 150 µg m^{-3} with a step of 10 µg m^{-3}, and a special threshold of 25 µg m^{-3} were used.

Total time of exceedance for the annual data sets signifies the number of days when the daily average concentration was greater than the threshold T. Because there were in total 15 concentration thresholds the TE_T (days) was performed for each one of those thresholds. In the end for each threshold a different number of days for the TE_T was calculated. Fig. 10 and Fig. 11 represent a visualization of the TE_T for Centar and Milano Pascal respectively.

Number of events for the annual data sets signifies the number of times when a threshold is continuously exceeded. There were again 16 thresholds and NE_T(-) was performed for each one. The results can be seen on the Fig. 10 and Fig. 11 which represent a visualization of the NE_T for Centar and Milano Pascal respectively.

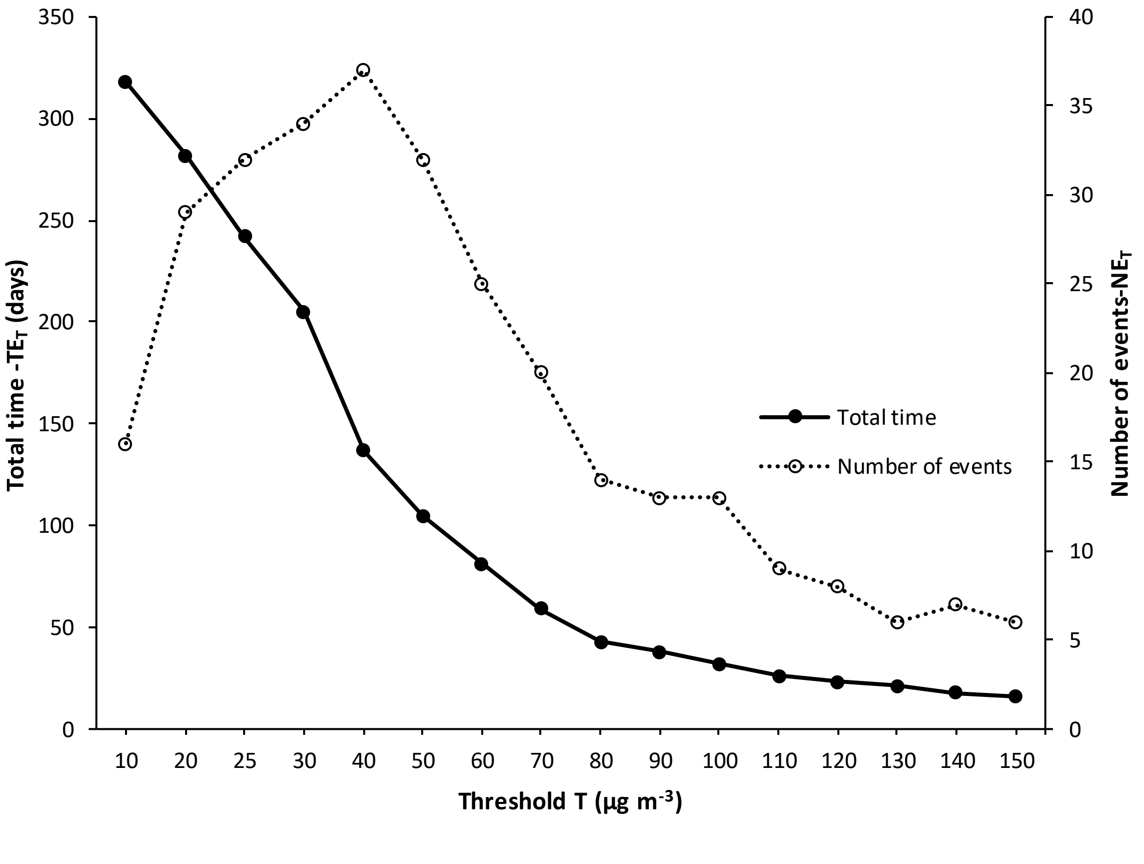


Fig. 10. Total time TE_T (days) and number of events NE_T (-) observed at the site of Centar during 2012.

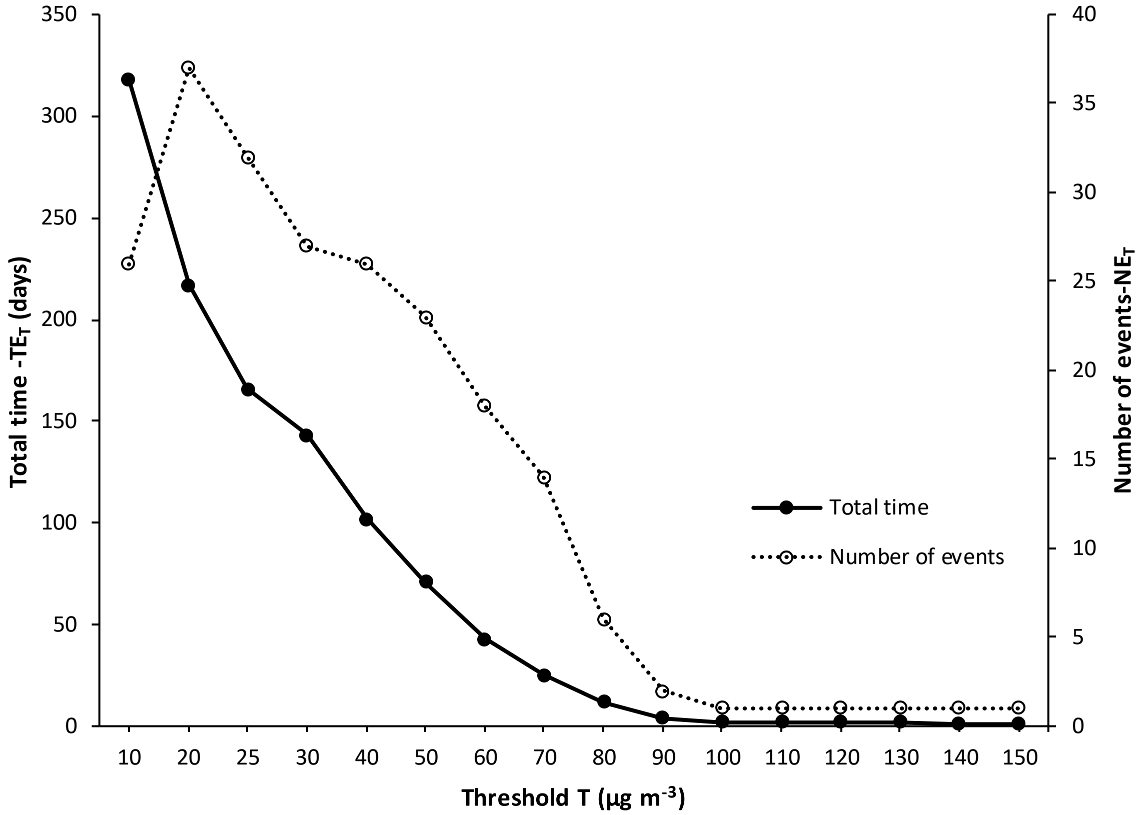


Fig. 11. Total time TE_T (days) and number of events NE_T (-) observed at the site of Milano Pascal during 2015.

If one looks closely at the two figures, representing Centar and Milano Pascal one cannot discern any big differences.

For the TE_T is the same total time is measured at both of the stations for the first threshold, then a decrease is noticed in the TE_T carried out until the last threshold where the TE_T for Centar is 16 days and 1 day for Milano Pascal.

For the NE_T the same extreme value is measured by both monitoring stations but not at the same threshold. NE_T curves follow the same pattern, and the differences are in the slightly bigger numbers for Centar.

On Figs. 10 & 11 there is a representation of only two stations, however if one pools all the data concerning the Total Time and the Number of events then one could represent the numerical data for those parameters for all the stations. In the Tables 4 & 5 the summary of the total time and number of events per threshold for all the stations is shown.

When looking closely at these tables one could notice the exact pattern seen for the two stations (Centar and Pascal), where the value of TE_T first threshold (10 µg m^{-3}) is lower than the value for the second threshold (20 µg m^{-3}), appears as a rightly skewed bell-shaped curve.

Table 4 Summary of Total time per threshold for all the stations

Summary Total time per threshold	Threshold																	
	10	20	25	30	40	50	60	70	80	90	100	110	120	130	140	150	160	170
Average	273.9	167.1	128.6	99.9	62.3	39.6	25.6	16.4	11.5	9.4	8.6	7.7	7.9	8.3	7.6	7.1	1.5	1
Standard Deviation	55.5	53.6	47	41.1	29.8	23.5	19.8	16.8	14.5	13.4	12.6	10.9	10.3	9.4	8.02	7.2	0.7	/
Min	107	73	46	31	22	11	2	2	1	1	1	1	1	1	1	1	1	1
Max	362	296	242	205	138	105	81	70	65	58	50	39	33	28	22	22	2	1
First quartile	253	129.5	92	68.5	38.5	19	12	6	3	2	1.3	1	1	1.3	1	1.3	1.3	1
Third quartile	312.5	197.5	153	116.5	77.5	49.5	32	18	12	8.8	5.8	7	14	16.5	14	11.3	1.8	1

Table 5 Summary of Number of events per threshold for all the stations

Summary Number of events per threshold	Threshold																	
	10	20	25	30	40	50	60	70	80	90	100	110	120	130	140	150	160	170
Average	27.6	33.1	26.3	25.6	19.3	13.8	10.5	7.5	5.1	4.1	3.9	3.5	3.5	3.5	3.7	3.4	1.5	1
Standard Deviation	10.9	6.7	6.7	6.5	7.1	6.3	5.7	5.	4.1	4.1	4.1	3.9	3.8	3.2	3	2.7	0.7	/
Min	1	13	16	11	7	4	2	1	1	1	1	1	1	1	1	1	1	1
Max	43	44	45	44	42	32	27	24	17	16	16	16	12	10	9	9	2	1
First quartile	21	30	24	22	14.5	9	6	4	2	1	1	1	1	1	1	1	1.3	1
Third quartile	36	37	31	28.5	22.5	16.5	13.5	8.8	6	5.8	4	3	5.5	5.8	6.3	5.8	1.8	1

The Fig. 12 bellow shows a plot of the average observed values for the Total time and the Number of events for each threshold of all the stations. It can be noticed right away that the Figs. 10 and 11 are follow the same pattern as the plot seen on Fig. 12, in other words they are not extreme.

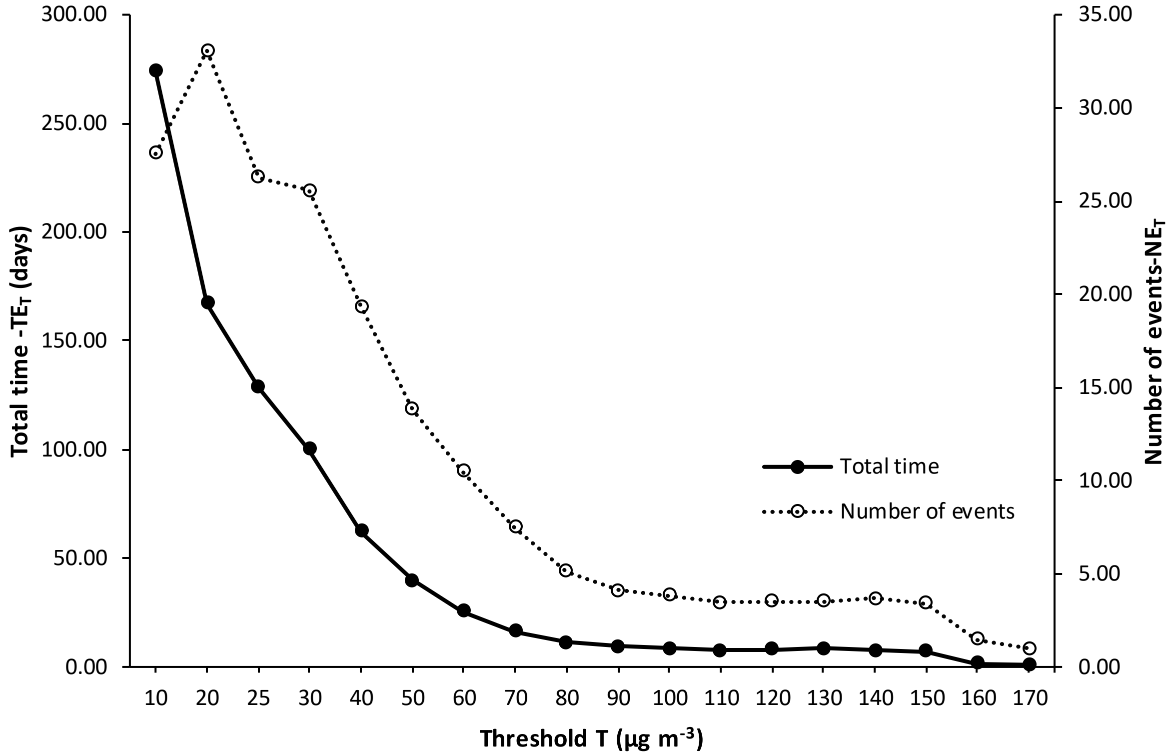


Fig. 12. The average observed values of the Total time TE_T (days) and number of events NE_T (-) for each threshold of all the stations.

2.3 Distribution of the events' duration

The distribution of the events' duration in addition to the NE_T shows how long the events for each of the 10 thresholds lasted. Along with the distribution of events' duration its most basic statistics (minimum, maximum, average and the interquartile range) will be shown. Fig. 13 and Fig. 14 represent a visualization of the distribution of the events' duration for Centar and Milano Pascal respectively.

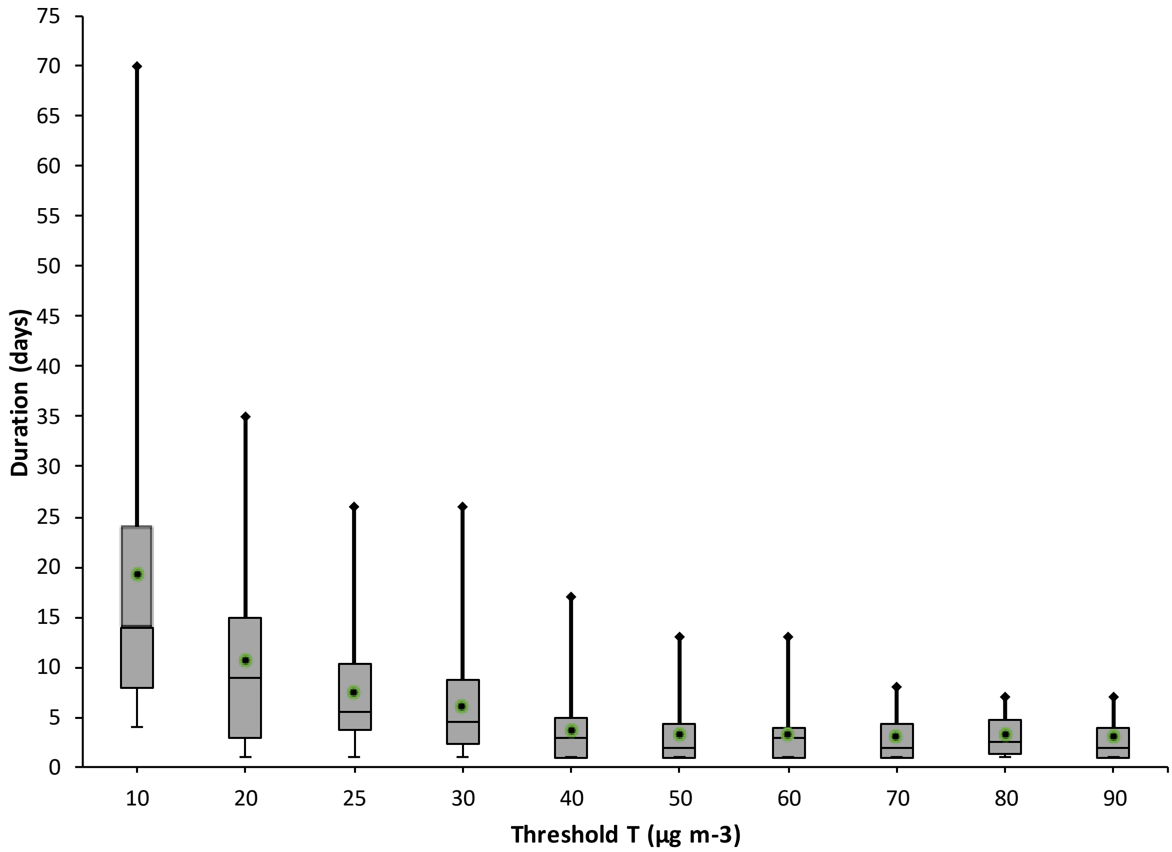


Fig. 13. Distributions of the events durations observed at the Centar site during 2012 (average: black dot; median: dash; IQR: gray box; minimum: small dash and maximum: diamond).

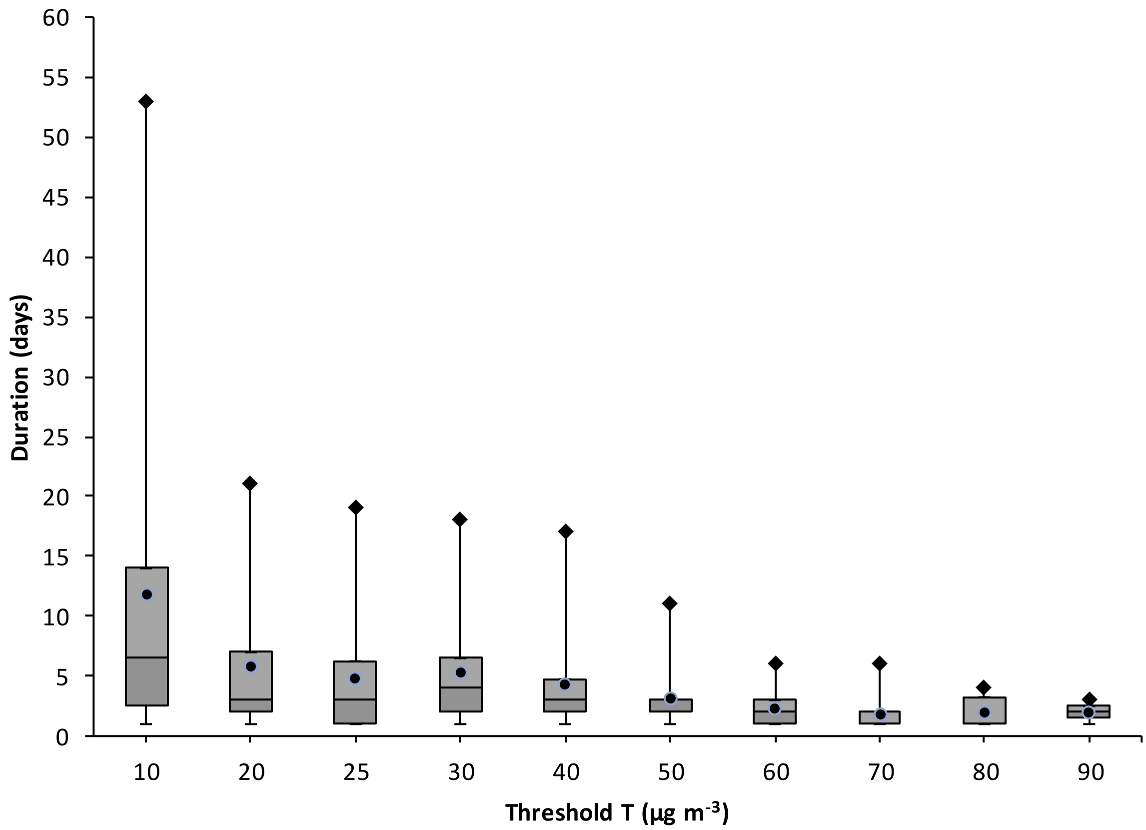


Fig. 14. Distributions of the events durations observed at the Milano Pascal site during 2015 (average: black dot; median: dash; IQR: gray box; minimum: small dash and maximum: diamond).

By looking at the TE_T, NE_T and the distributions of the events' duration, one could conclude that lowest thresholds are being exceeded for the longest, accounting for the highest duration of events. Whereas for the highest thresholds one can observe that the TE_T, NE_T are decreasing. Resulting in low TE_T, NE_T with events lasting for a short amount of time.
From the Fig. 13 and Fig. 14 it can be concluded that the data is skewed to the right since the mean value is bigger than the median.

If one looks closely at the Fig. 13 and Fig. 14, one can see that there are differences in the extreme durations most notably for the first 2 thresholds, where the longest event in Centar lasted 70 days and the longest event in Milano Pascal lasted only 55. For the other thresholds, the differences are minimal.

2.4 Theoretical model for distribution of duration

In this paragraph, a theoretical model will be fitted to the duration of the events for the following thresholds: 10, 20, 25, 30 µg m^{-3} (Fig. 15 and Fig. 16). This case-study uses PM2.5 concentrations and the EU concentration limits for these particles, as annual average is 25 µg m^{-3}, for this exact reason, the chosen thresholds are exactly those. As mentioned before, the data of the duration of the events is skewed to the right, indicating the arithmetic mean is bigger than the median of this data. Suggested by some literature (Giugliano, 1998); (Georgopoulos, 1982)) the two-parameter log-normal distribution model was being used for fitting the duration of events data for the four thresholds. The fitting of the data was evaluated by Kolmogorov-Smirnov goodness of fit test at 5% significance. Although the test was negative, meaning that the data was not log-normally distributed, the problem can be traced to high number of short events (G. Lonati, 2011), for example if one removes some of the short events the Kolmogorov-Smirnov test becomes positive. Other than this the log-normal distribution fits rather well the distribution data.

For further inspection of how well the data was fitted one can use the performance parameters FB (fractional bias), FE (fractional error) and R (correlation coefficient). They are represented by following two expressions:

$$FB = 2\,\frac{(Pi - Oi)}{(Pi + Oi)} \qquad \text{Eq. 1}$$

$$FE = 2\,\frac{|(Pi - Oi)|}{|(Pi + Oi)|} \qquad \text{Eq. 2}$$

Where P is predicted, O is observed values. For predictions, as close to the observed value as possible, P=O the fractional bias and error are equal to zero, FB=0=FE.

The percentiles of the predictions of the durations of the events exceeding a threshold T computed by the lognormal model are plotted versus the percentiles of the observed durations exceeding the same threshold T. This case-study considers only 4 concentration thresholds: 10, 20, 25 and 30 µg m^{-3}. From Figs. 17 through 20 one could appreciate the predicted versus the observed percentiles of the durations events for these thresholds.

Summary statistics are also available as seen on Table 6:

- The correlation coefficirent for all the percentiles is excellent, ranging from 83-98% for 50^{th} percentiles, 88-94% for 75^{th} percentiles, 90-94% for 80^{th} percentiles, 85-96% for 90^{th} percentiles, 85-88% for 95^{th} percentiles, 68-82% for 99^{th} percentiles.
- Fractional Bias is around 0 which signifies that the predicted values are very close to the observed ones: 0.04-0.14 for 50^{th} percentiles, -0.02-0.07 for 75^{th} percentiles, 0-0.04 for 80^{th} percentiles, -0.08-0.03 for 90^{th} percentiles, -0.04-0.15 for 95^{th} percentiles, 0.08-0.43 for 99^{th} percentiles.

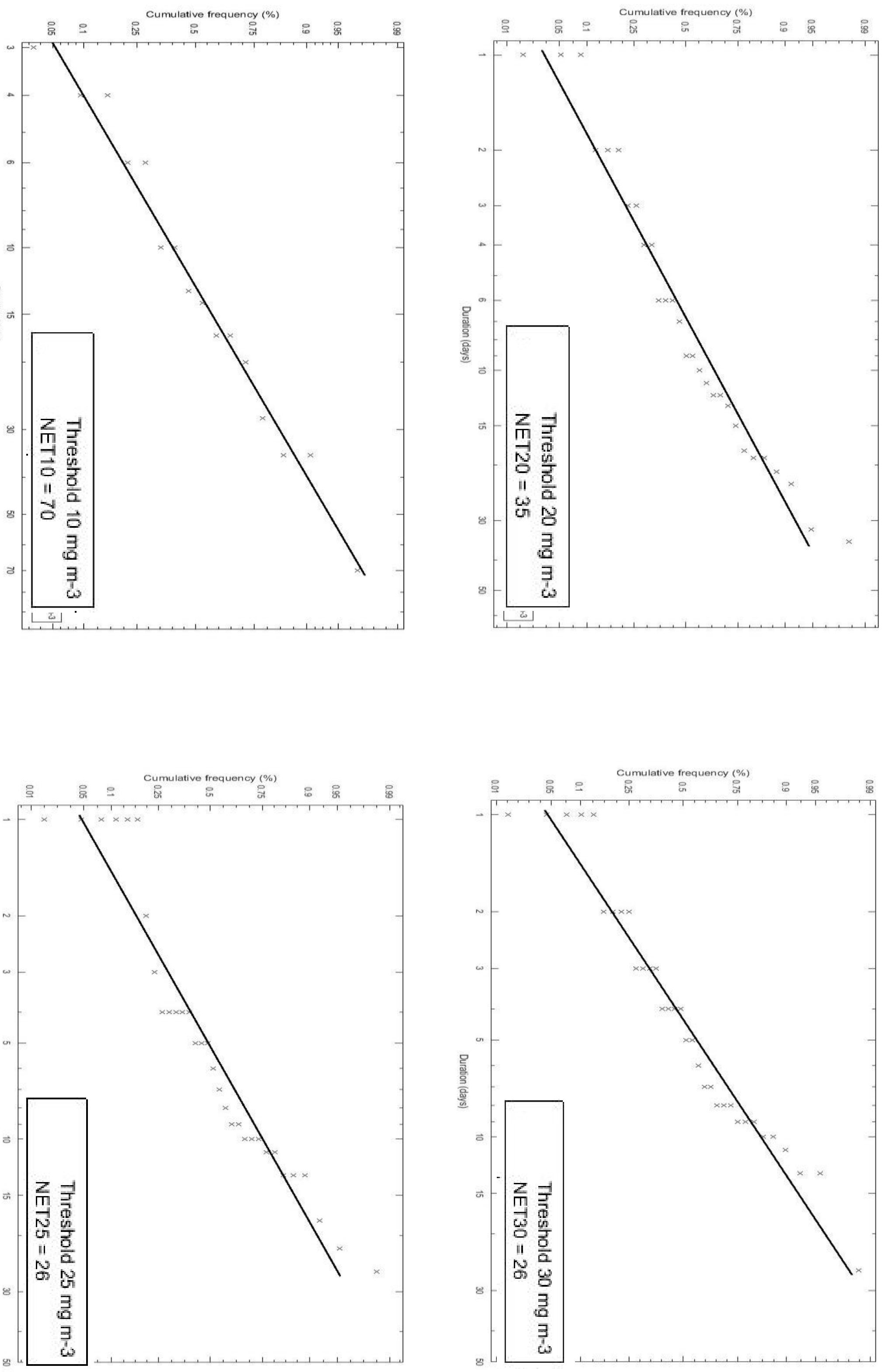


Fig. 15. Distributions of the durations observed at Centar site in 2012 (circles), fitted lognormal distributions (thick line) for given thresholds: for PM pollution (10, 20, 25, 30 µg m^{-3}))

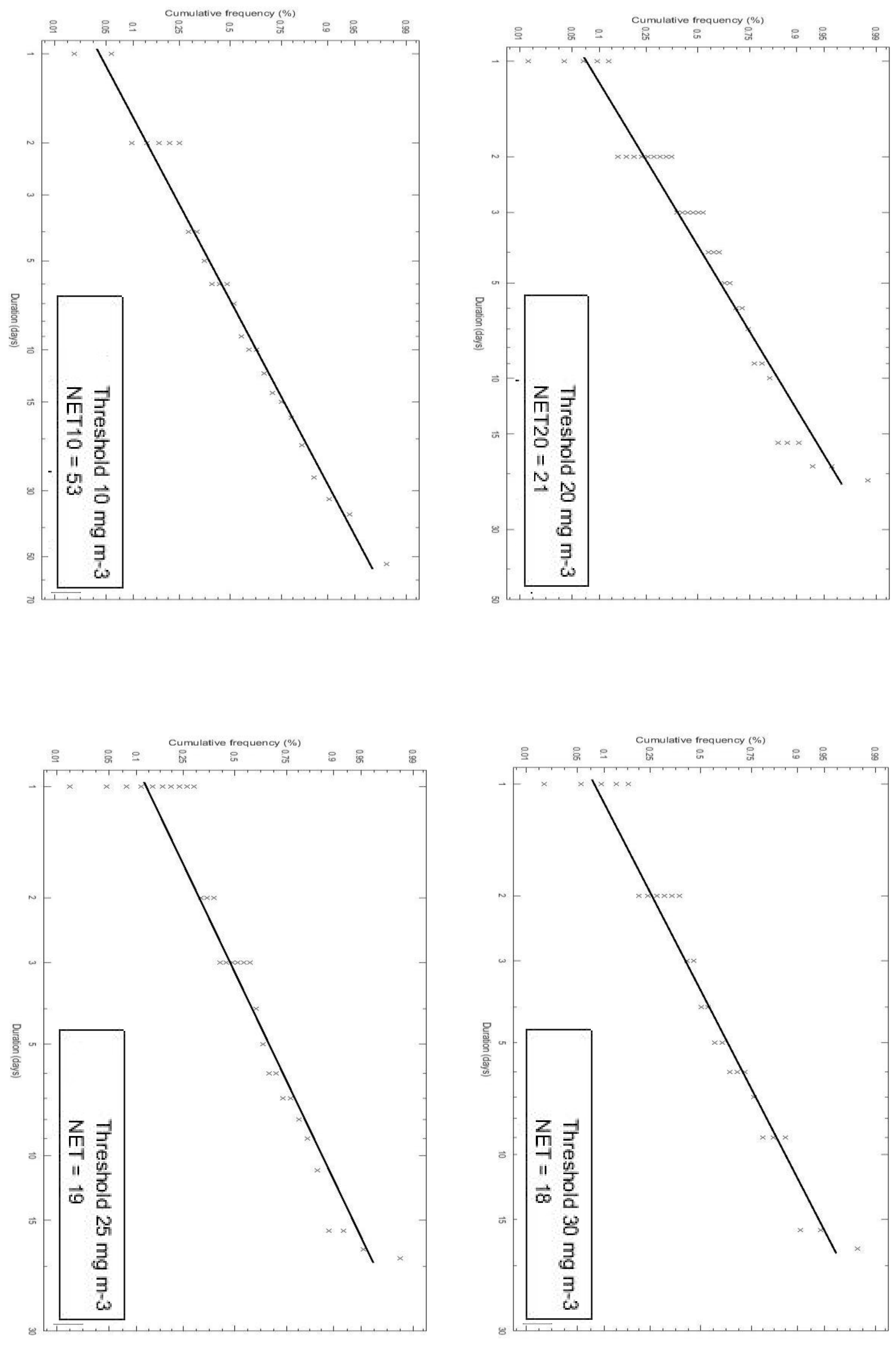


Fig. 16. Distributions of the durations observed at Milano Pascal site in 2015 (circles), fitted lognormal distributions (thick line) for given thresholds: for PM pollution (10, 20, 25, 30 µg m^{-3}))

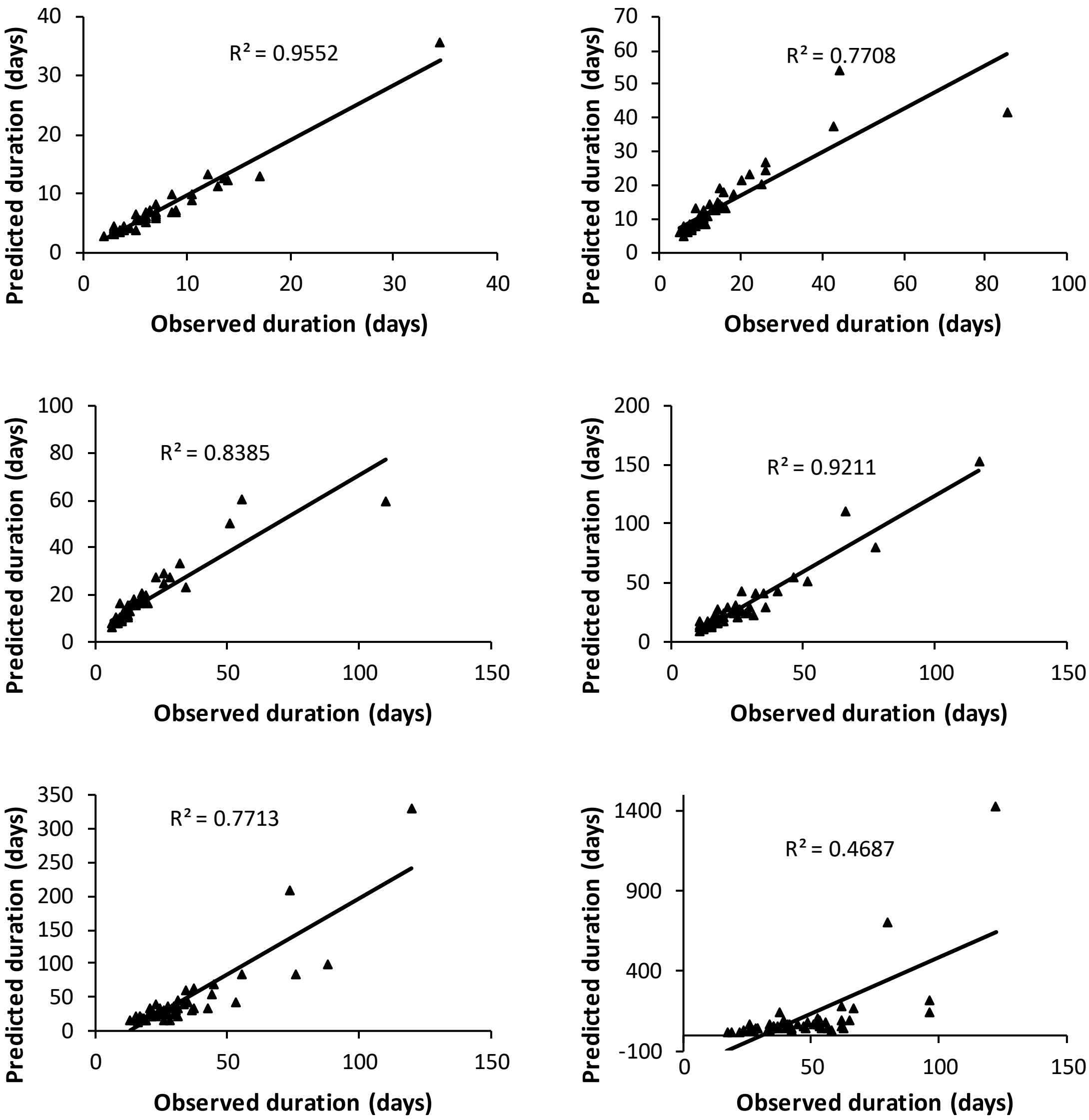


Fig. 17. Duration of the events exceeding the 10 µg m^{-3} threshold: observed and predicted percentiles by the lognormal models for all the stations (50th percentile, upper left; 75th percentile, upper right; 80th percentile, middle left; 90th percentile, middle right; 95th percentile, lower left; 99th percentile, lower right).

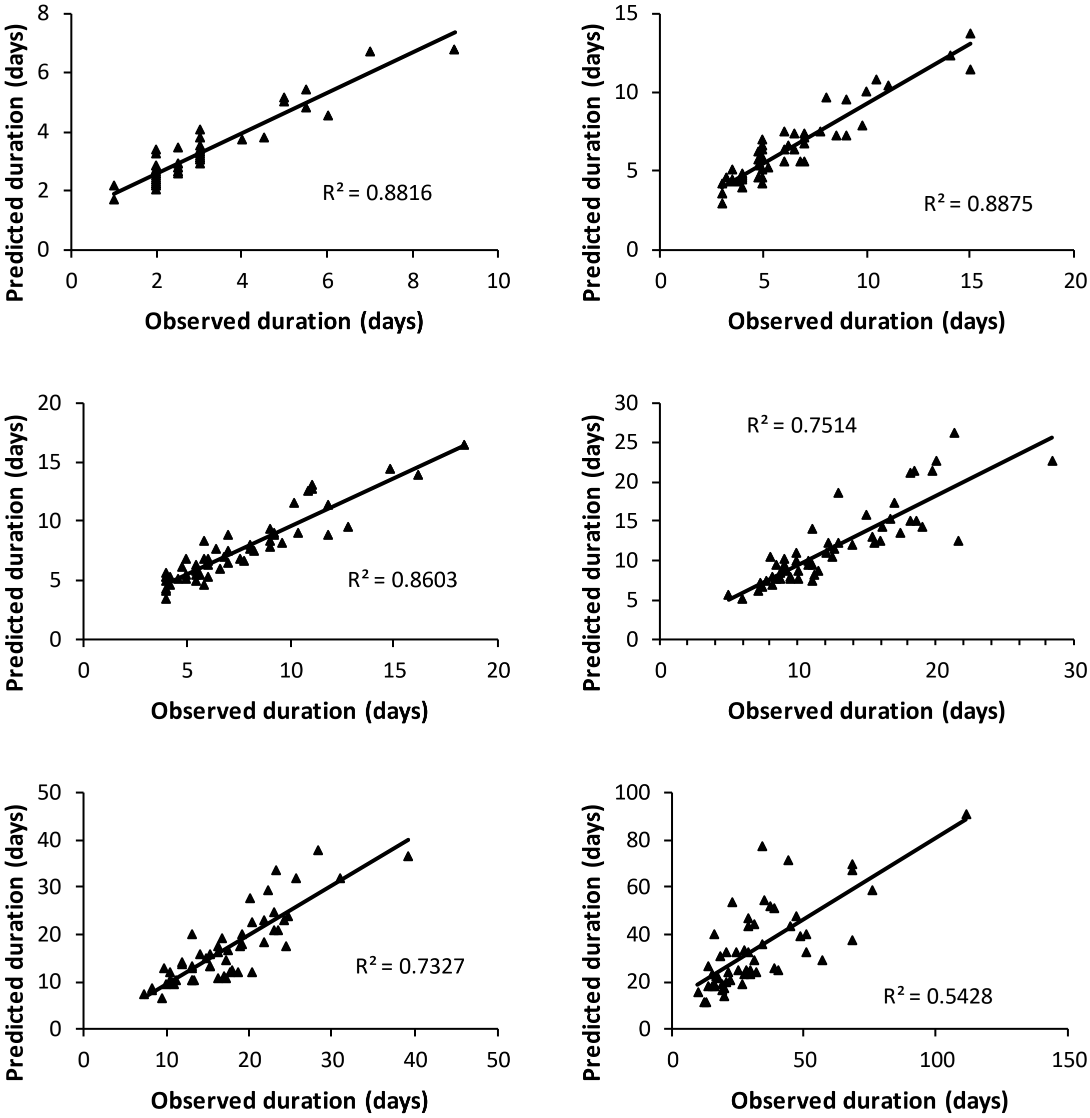


Fig. 18. Duration of the events exceeding the 20 µg m^{-3} threshold: observed and predicted percentiles by the lognormal models for all the stations (50th percentile, upper left; 75th percentile, upper right; 80th percentile, middle left; 90th percentile, middle right; 95th percentile, lower left; 99th percentile, lower right).

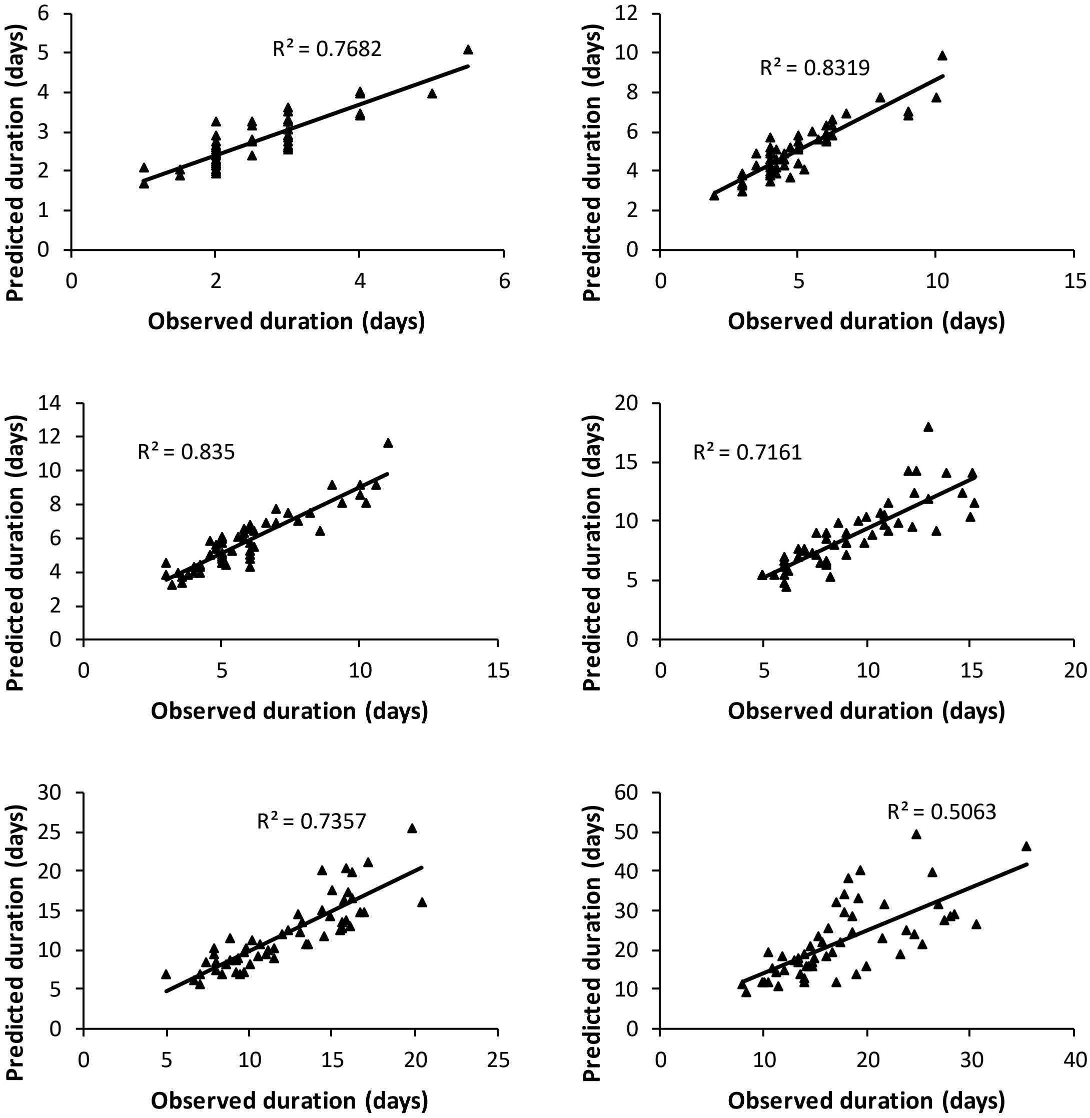


Fig. 19. Duration of the events exceeding the 25 μg m^{-3} threshold: observed and predicted percentiles by the lognormal models for all the stations (50th percentile, upper left; 75th percentile, upper right; 80th percentile, middle left; 90th percentile, middle right; 95th percentile, lower left; 99th percentile, lower right).

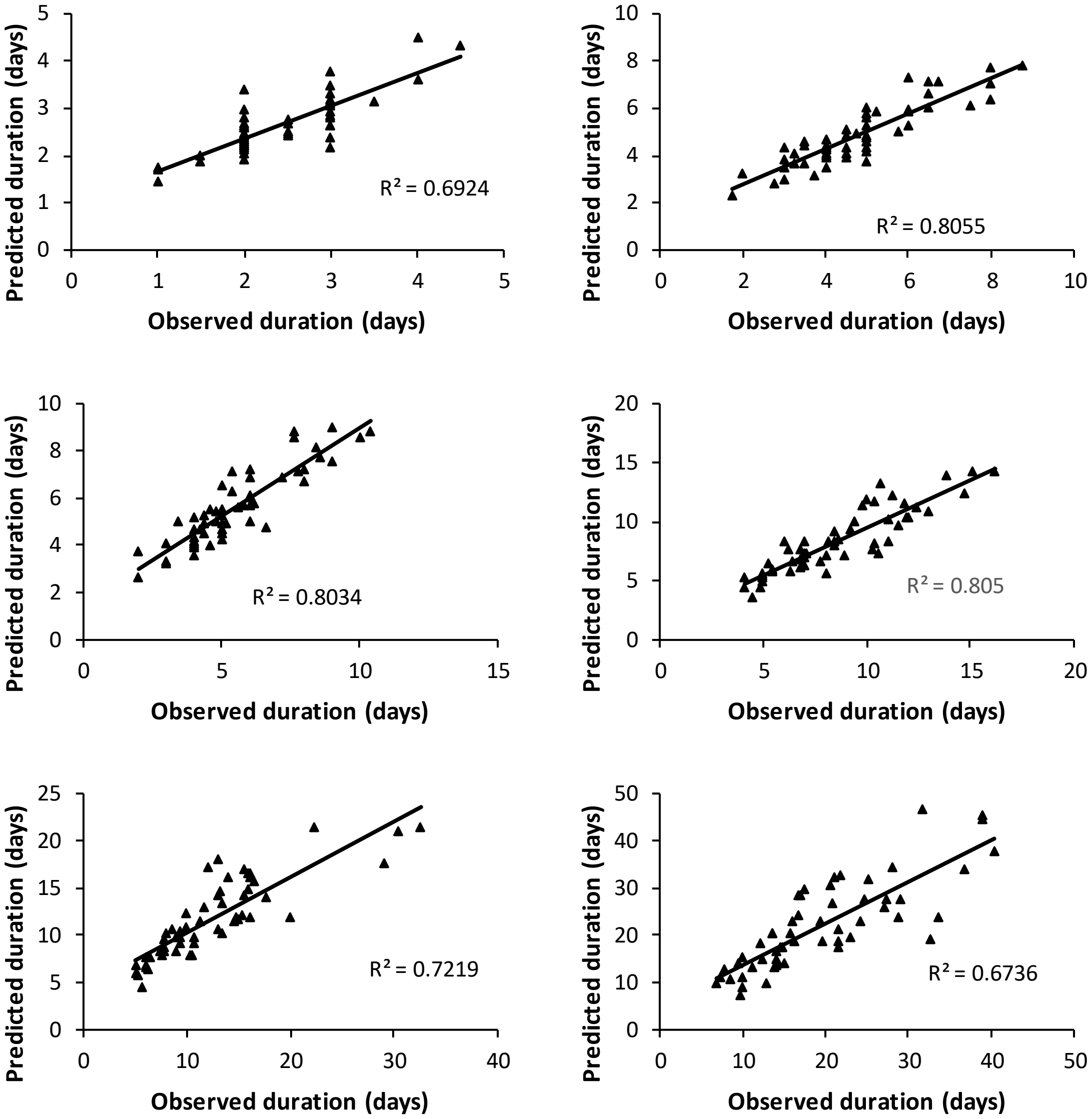


Fig. 20. Duration of the events exceeding the 30 µg m^{-3} threshold: observed and predicted percentiles by the lognormal models for all the stations (50th percentile, upper left; 75th percentile, upper right; 80th percentile, middle left; 90th percentile, middle right; 95th percentile, lower left; 99th percentile, lower right).

Table 6 Average values of the performance parameters (FB, FE, R) between the predicted values by the lognormal model and the observed percentiles of the duration exceeding thresholds of 10, 20, 25 and 30 μg m^{-3}.

Percentile	Threshold 10 μg m^{-3}						Threshold 20 μg m^{-3}					
	0.50	0.75	0.80	0.90	0.95	0.99	0.50	0.75	0.80	0.90	0.95	0.99
FB	0.04	-0.02	0.01	0.03	0.15	0.43	0.14	0.07	0.03	-0.08	-0.04	0.08
FE	0.14	0.12	0.11	0.15	0.25	0.48	0.18	0.14	0.13	0.16	0.17	0.28
R	0.98	0.88	0.92	0.96	0.88	0.68	0.94	0.94	0.93	0.87	0.86	0.74
Percentile	Threshold 25 μg m^{-3}						Threshold 30 μg m^{-3}					
	0.50	0.75	0.80	0.90	0.95	0.99	0.50	0.75	0.80	0.90	0.95	0.99
FB	0.12	0.04	0.00	-0.05	-0.02	0.21	0.12	0.04	0.04	0.02	-0.02	0.13
FE	0.17	0.12	0.11	0.13	0.14	0.27	0.16	0.13	0.13	0.16	0.17	0.23
R	0.88	0.91	0.91	0.85	0.86	0.71	0.83	0.90	0.90	0.90	0.85	0.82

3 Results and discussion

3.1 Model for the total time of exceedance

In order for one to derive the model for the total time of exceedance one must pool all the data available (concentration of PM2.5), from all of the stations for every year from 2012-2017. TE_T represents the total time, in this case-study days, because the gathered data is average daily concentrations of PM2.5, when a fixed concentration threshold is exceeded.
As seen previously all the data sets share the same analogy when it comes to the total time of exceedance - TE_T. If one sees Fig. 10 and Fig. 11 one can conclude that, as the threshold increases the total time of exceedance decreases. For example, for a threshold of 10 µg m^{-3} the value of TE_T is around 318 days, which is almost all the days of the year, considering the missing data, and for a threshold of 150 µg m^{-3} the TE_T has drastically decreased to 1 through 15 days of the year.
Following is the parameterized empirical model of the annual data for total time of exceedance (TE_T), made to explain the observed data, and can be further used to estimate concentration levels in future scenarios. This model is a function of R_{TC} = T/Ca, between T threshold and the annual average concentration Ca. TE_T was calculated for all 55-data series, by using the 16 thresholds ranging from 10 µg m^{-3} to 150 µg m^{-3} with a step of 10 µg m^{-3}, and a special threshold of 25 µg m^{-3}. Because there were in total 16 thresholds the TE_T was performed for each one of those thresholds. At the end, the TE_T data was plotted on the y-axis and R_{TC} on the x-axis, as seen on Fig. 21. The data was rather well fitted with the following exponential expression:

$$TE_T = a_1 \exp\left(b_1 R_{TC}\right) \qquad \text{Eq. 3}$$

The parameters a_1 = 441.9, b_1 = -1.319, shown in the eq. 3 were calculated by least squares. In order to achieve the decrease of the TE_T as the threshold also decreases the value of the parameter b_1 had to be negative.
So as to examine the accuracy of the parameterized empirical model for the total time of exceedance TE_T eq.3, a certain evaluation was needed. As mentioned before one way to determine the accuracy is by using the statistical parameters eq. 1, and eq. 2. Because of the many data sets, an average value of the FB and FE was needed. To do so one needs to calculate the FB and FE for each concentration threshold of every data series. The summary statistics can be seen on Table 7, where one could see that the average value of FB and FE is around 0, meaning that the prediction was almost perfect. The distributions of the FB are presented on Fig. 22, there one could see that the average value of the FB is around 0 with small fluctuations and the values range from 1.81 to -1.79. For lower thresholds, the model is more accurate, but also for the higher thresholds the extreme values for the FB are acceptable.

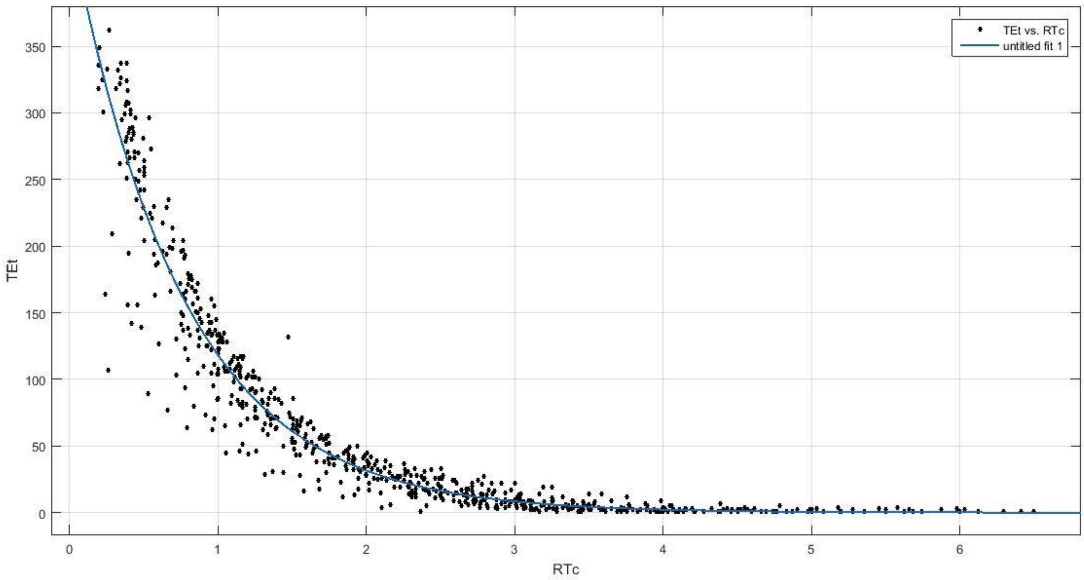


Fig. 21. Empirical Model, TE_T as a function of $R_{TC} = T/Ca$, T - threshold, Ca – annual average concentration.

Table 7 Summary statistics of the performance parameters FB and FE for eq. 3.

Parameter TE_T equation 3	FB	FE
Average	0.00	0.35
Standard deviation	0.51	0.37
Minimum	-1.79	0.00
Maximum	1.81	1.81
First quartile	-0.19	0.09
Third quartile	0.22	0.51

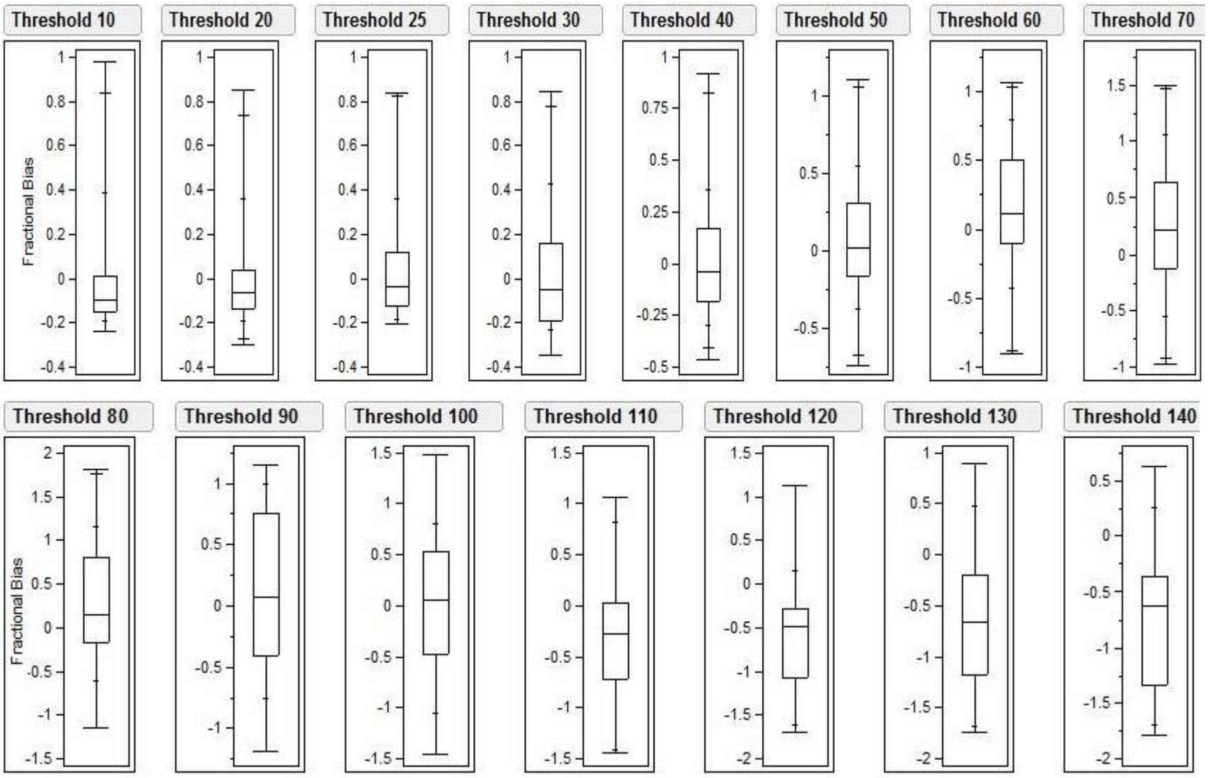


Fig. 22. Distributions of FB values for TE_T description by means of eq. (4) (average: black dot; IQR: gray box; min. and max. value: whiskers). (the scale of each boxplot is assigned to the right of the boxplot)

3.2. Model for the number of events

In order for one to derive the model for the number of events one must pool all the data available (concentration of PM2.5), from all of the stations for every year from 2012-2017. NE_T represents the number of events of exceedance, a dimensionless variable, describing how many events exceed a fixed concentration threshold. Where an event is characterized by a series of days when the fixed threshold is exceeded without interruption.

In Fig. 10 and Fig. 11, the NE_T part, one could notice that both of those figures share a single pattern, this is true for all of the number of events NE_T curves. It is clear that every time series starts with low number of events for the first, and then at the second concentration threshold there is an increase of the number of events and after that a decrease. Actually, it is quite logical that the number of events should be few when the threshold is low, because those events last a lot longer than the events for the higher concentration thresholds, as can be seen by the distribution of duration on Fig. 13 and Fig. 14. What this means is that the development of the empirical model will be not as easy as with the TE_T which was a simple exponential model. Now one is facing a right skewed data that starts with low number of events then reaches a maximum at around 35 number of events, for the second or third concentration threshold, then a decrease reaching zero events for the highest concentration threshold.

If one observes closely the NE_T curves, one could right away notice that the data looks like a skewed to the right normal distribution. So, the wisest choice is to make empirical model of the number of events very similar to the expression for this kind of distribution.

Following is the parameterized empirical model of the yearly data for the number of events (NE_T) exceeding a threshold T, made to explain the observed data, and can be further used to estimate concentration levels in future scenarios. This model is a function of R_{TC} = T/Ca, between T threshold and the annual average concentration Ca. NE_T was calculated for all 55-data series, by using the 16 thresholds ranging from 10 µg m^{-3} to 150 µg m^{-3} with a step of 10 µg m^{-3}, and a special threshold of 25 µg m^{-3}. Because there were in total 16 thresholds the NE_T was performed for each one of those thresholds. At the end, the NE_T data was plotted on the y-axis and R_{TC} on the x-axis, as seen on Fig. 23. The data was rather well fitted with the following expression:

$$NE_T = 365 \{a_2 + b_2 \exp [-0.5 (\ln (R_{TC} / c_2) / d_2)^2]\} \qquad \text{Eq. 4}$$

The parameters a_2 = -0.00073, b_2 = 0.09294, c_2 = 0.6364, d_2 = 0.8097, shown in the eq. 3 were calculated by least squares.

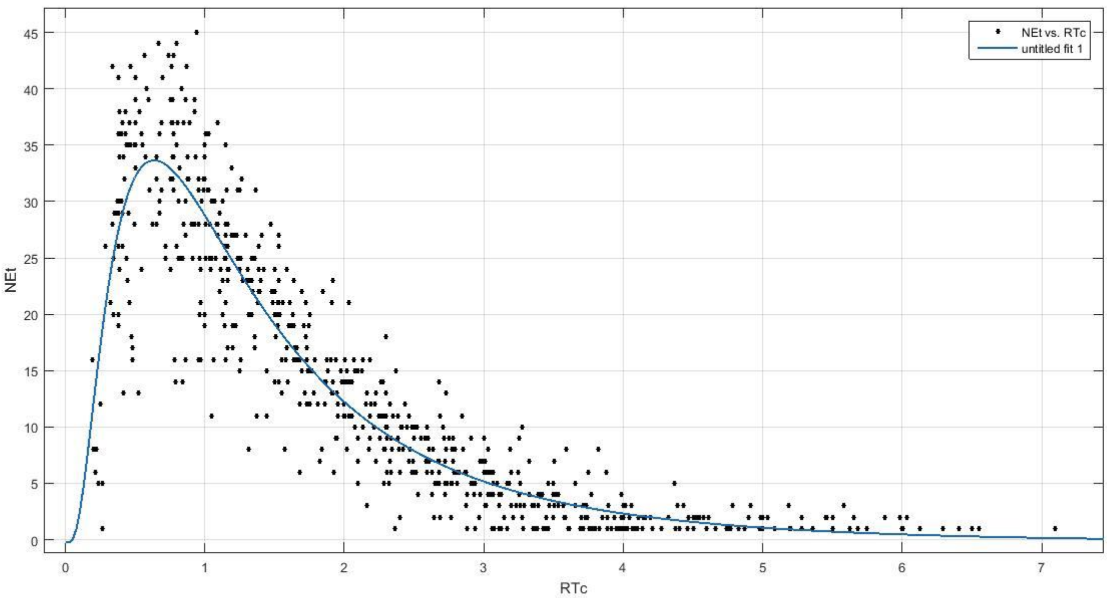


Fig. 23. Empirical Model, NE_T as a function of R_{TC} = T/Ca, T - threshold, Ca – annual average concentration.

So as to examine the accuracy of the parameterized empirical model for the number of events eq. 4, a certain evaluation was needed. As mentioned before one way to determine the accuracy is by using the statistical parameters eq. 1, and eq. 2. Because of the many data sets, an average value of the FB and FE was needed. To do so one needs to calculate the FB and FE for each concentration threshold of every data series. The summary statistics can be seen on Table 8, where one could see that the average value of FB and FE is around 0.3, meaning that the prediction was not so perfect. The distributions of the FB are presented on Fig. 24, there one could see that the average value of the FB is around 0.3 with small fluctuations and the values range from -0.6 to 1.86.

Table 8 Summary statistics of the performance parameters FB and FE for eq. 4.

Parameter NE_T equation 4	FB	FE
Average	0.24	0.37
Standard deviation	0.46	0.37
Minimum	-0.60	0.00
Maximum	1.86	1.86
First quartile	-0.09	0.10
Third quartile	0.49	0.50

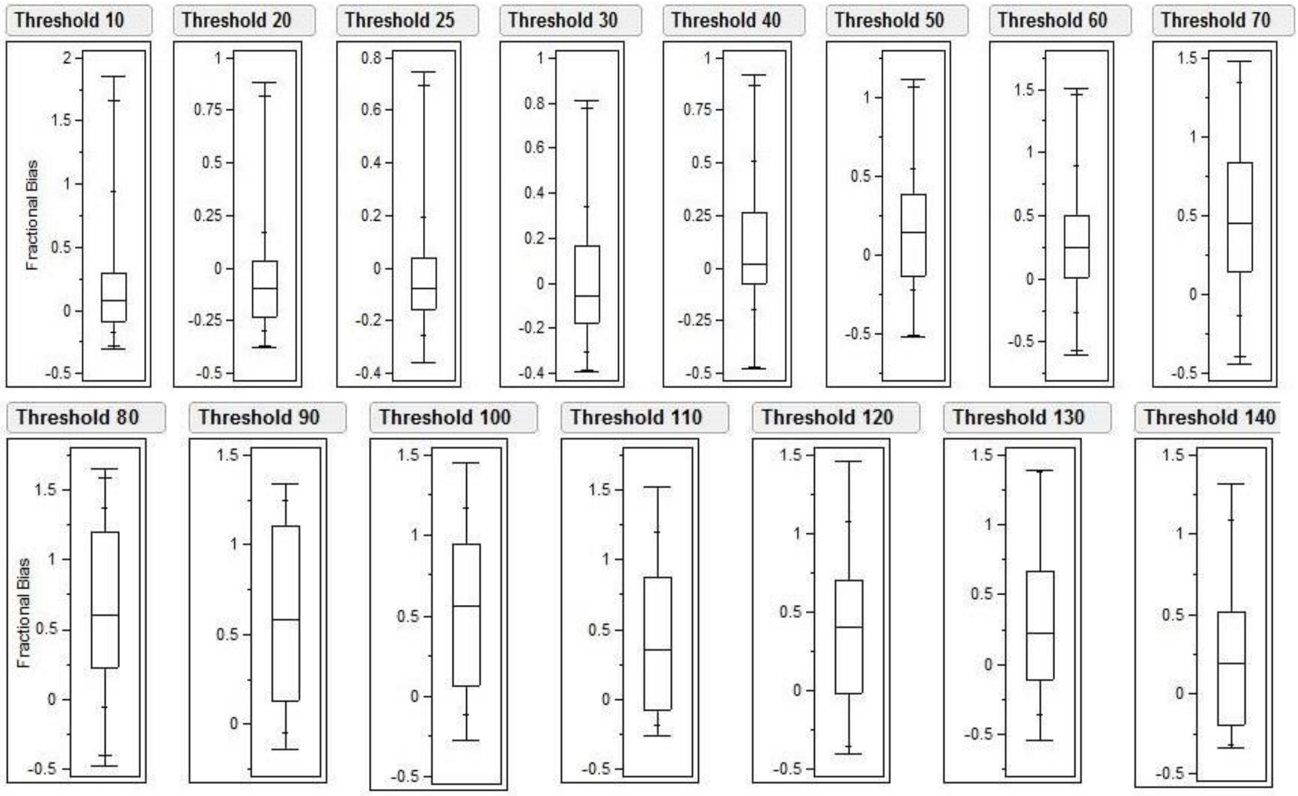


Fig. 24. Distributions of FB values for NE_T by using eq.4 (average: line inside the box; IQR: gray box; min. and max. value: whiskers). (the scale of each boxplot is assigned to the right of the boxplot)

3.3 Model for the distribution of duration

For the distribution of the duration a lognormal model will be used, because it was seen already that it fits well the data (G. Lonati, 2011). For the following model as well as the other ones before all the 55-data series were used. This time however, a two parameters lognormal distribution will be calculated for each of the 4 thresholds ranging from 10 µg m^{-3} to 30 µg m^{-3} (10, 20, 25 and 30) for every one of the data series, resulting in σ_T and μ_T, for each threshold of all the time series. At the end, they are organized and plotted as a function of R_{TC}, as seen on Figs. 25 and 26. The formulations for σ_T and μ_T are derived by linear regression of the said figures:

$$\sigma_T = -0.244\, R_{TC} + 1.156 \qquad \text{Eq. 5}$$

$$\mu_T = -0.908\, R_{TC} + 1.926 \qquad \text{Eq. 6}$$

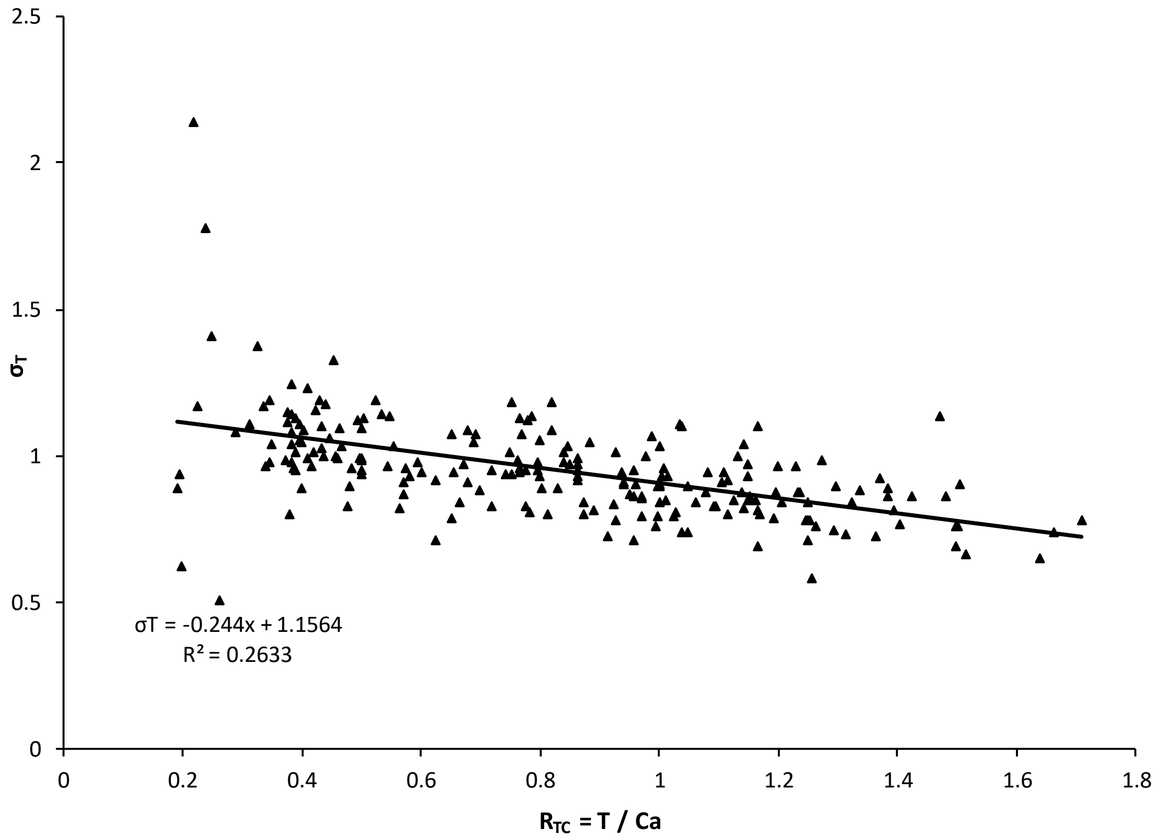


Fig. 25. Linear regression for the parameter (σ_T) of the log-normal model for the distribution of the duration of events that exceed a threshold T as a function of R_{TC} = T/Ca, T - threshold, Ca – annual average concentration.

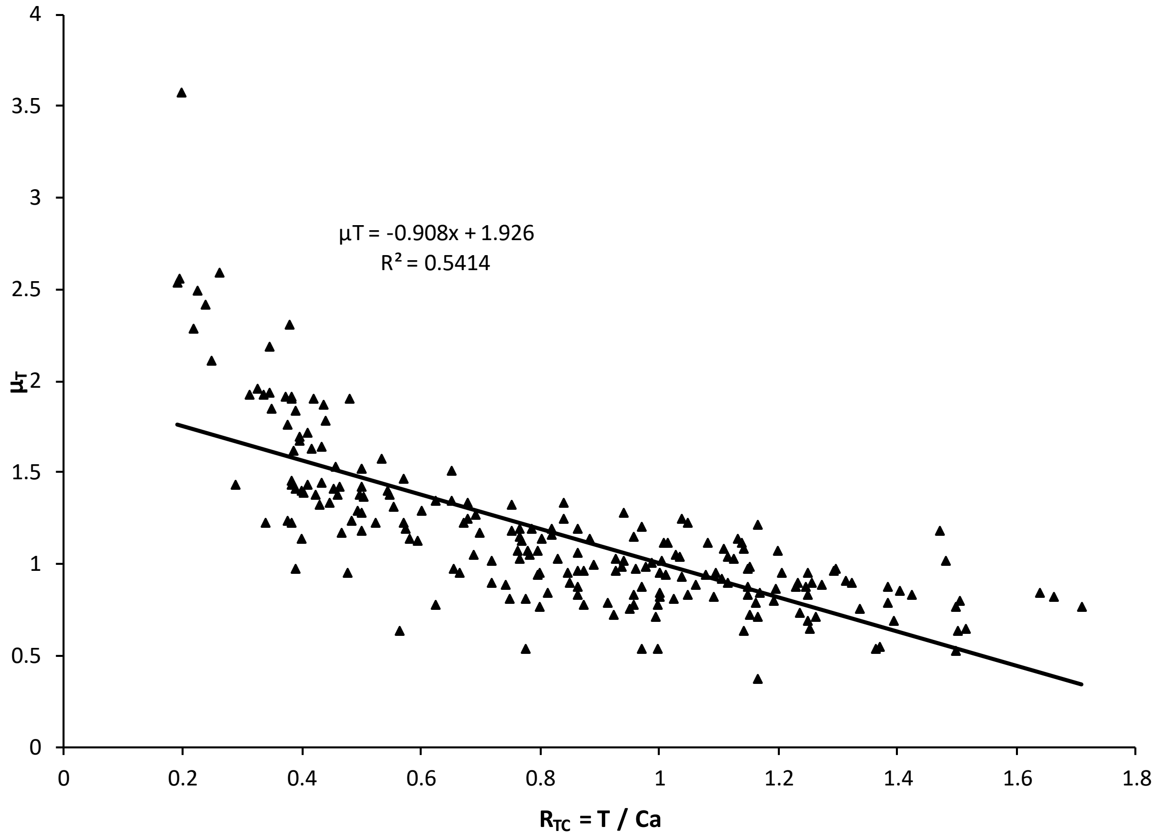


Fig. 26. Linear regression for the parameter (μ_T) of the log-normal model for the distribution of the duration of events that exceed a threshold T as a function of R_{TC} = T/Ca, T - threshold, Ca – annual average concentration.

As one could notice the lognormal parameters are inversely proportional with the R_{TC}, they decrease while R_{TC} increases. R_{TC} decreases as the Ca – average annual concentration increases in turn the lognormal parameters increase thus we have duration of events over a given threshold lasting longer.

$$F(D_T) = \int_0^D f_T\, dx = \int_0^D \frac{1}{x\sigma_T\sqrt{2\pi}} \exp\{-\frac{1}{2}[\frac{\ln(x)-\mu_T}{\sigma_T}]^2\} \quad \text{Eq. 7}$$

By combining eqs. 5, 6 and the lognormal distribution one could predict the distribution of duration D_T, if only one knows the Ca. Through eq. 8 and eq. 9 one can calculate the arithmetic mean and mode of the lognormal distribution.

$$\text{Arithmetic mean} = \exp(\mu + 0.5\sigma^2) \quad \text{Eq. 8}$$

$$\text{Mode} = \exp(\mu - \sigma) \quad \text{Eq. 9}$$

By plugging in eq. 5 and 6, for μ and σ, one could estimate the average duration μ_{DT} (days) of the events of exceedance (eq.10) and the mode M_{DT} i.e. the most frequently occurred duration (eq.11). Because these parameters are a function of R_{TC} = T/Ca, between T threshold and the annual average concentration Ca, they are dependent on the ratio between the T and the Ca.

$$\mu_{DT} = \exp(0.059\, R_{TC}^2 - 1.19\, R_{TC} + 3.26) \quad \text{Eq. 10}$$

$$M_{DT} = \exp(-1.152\, R_{TC} + 3.082) \quad \text{Eq. 11}$$

So as to examine the accuracy of the parameterized empirical model for the distribution of the duration eqs. 5, 6 and 7, a certain evaluation was needed. As mentioned before one way to determine the accuracy is by using the statistical parameters eqs. 1, 2. Because of the many data sets, an average value of the FB and FE and the correlation coefficient R was needed. To do so one needs to make a comparison between the predicted and observed highest percentiles (50th, 75th, 90th, 95th, 99th) of the duration distributions for all data series. The summary statistics can be seen on Table 9, which shows that the evaluations done by the eqs. 5, 6 and 7 are quite precise with the correlation ranging from 42-82%, and FB always around zero, which signifies that the predictions are very close to the observations.

Table 9 Average values of the performance parameters (FB, FE, R) between the predicted values by the lognormal model and the observed percentiles of the duration exceeding thresholds of 10, 20, 25 and 30 μg m^{-3}.

Percentile	Threshold 10 μg m^{-3}						Threshold 20 μg m^{-3}					
	0.50	0.75	0.80	0.90	0.95	0.99	0.50	0.75	0.80	0.90	0.95	0.99
FB	-0.11	-0.13	-0.13	-0.13	-0.13	-0.13	0.09	0.07	0.07	0.06	0.06	0.04
FE	0.31	0.32	0.32	0.33	0.34	0.38	0.17	0.19	0.20	0.22	0.24	0.29
R	0.67	0.76	0.75	0.66	0.57	0.45	0.79	0.82	0.82	0.81	0.80	0.77
Percentile	Threshold 25 μg m^{-3}						Threshold 30 μg m^{-3}					
	0.50	0.75	0.80	0.90	0.95	0.99	0.50	0.75	0.80	0.90	0.95	0.99
FB	0.11	0.15	0.16	0.09	0.20	0.23	-0.07	-0.08	-0.08	-0.09	-0.09	-0.10
FE	0.19	0.23	0.24	0.43	0.30	0.35	0.18	0.20	0.21	0.23	0.25	0.29
R	0.45	0.43	0.42	0.40	0.38	0.35	0.62	0.63	0.63	0.61	0.59	0.54

3.4. Model for the number of events of given duration

The last step of this report is the development of the model for the number of events of given duration. For this particular model, a combination of some of the already seen eqs. 4 and 7 is required. With the eq. 4 one could compute the annual Number of Events NE_T, whereas eq. 7 allows computation of the distribution of the duration. Put together these two equations and one could compute the cumulative number of events $NE_{D,T}$ that are exceeding a certain concentration threshold T for a given time lower than or equal to the certain duration D.

$$NE_{D,T} = NE_T F(DT) \quad \text{Eq. 12}$$

According to the Table 10, the summary statistics of the performance parameter FB (Fractional Bias) have been calculated. What they show is the following:

- The positive/negative value of the average FB tells us that the model is over-predicting/under-predicting the number of events
- The other statistics are used to make the distribution of the FB values seen on the Fig. 27 for a better visual understanding

For a concentration threshold of 10 µg m^{-3} the average FB value is positive resulting in over-prediction of the number of events, this can be appreciated on Fig. 28 and Fig. 29, on the other hand for the concentration threshold of 20, 25 and 30 µg m^{-3} a noticeable under-prediction can be seen.

Table 10 Summary statistics of the performance parameter FB for the number of events $NE_{D,T}$ calculated with eq. 1.

Parameter	Threshold			
	10 µg m^{-3}	20 µg m^{-3}	25 µg m^{-3}	30 µg m^{-3}
Average	0.24	-0.10	-0.14	-0.06
Standard deviation	0.54	0.31	0.34	0.28
Minimum	-1.49	-1.53	-1.43	-1.38
Maximum	2.00	1.10	0.76	0.87
First quartile	-0.10	-0.26	-0.32	-0.19
Third quartile	0.45	0.00	0.07	0.09

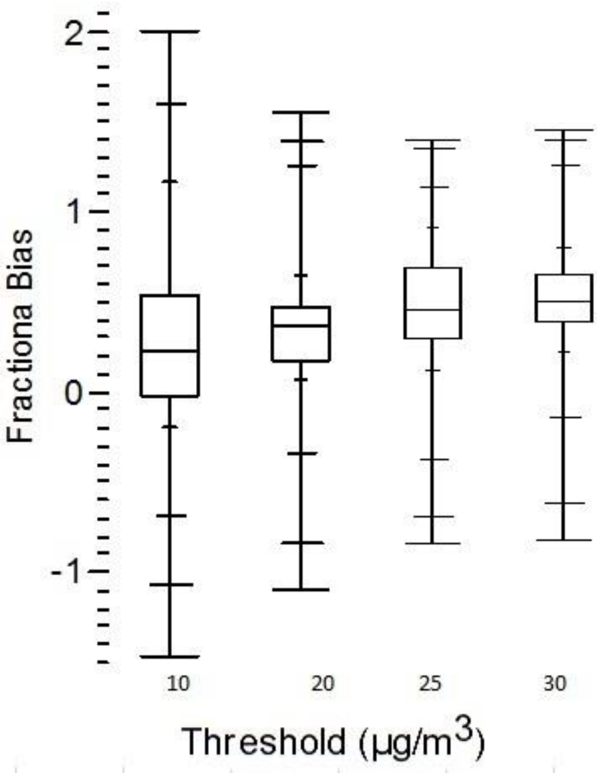


Fig. 27. Distributions of FB values for $NE_{D,T}$ eq. 12 for the thresholds of 10, 20, 25, 30 µg m^{-3}.

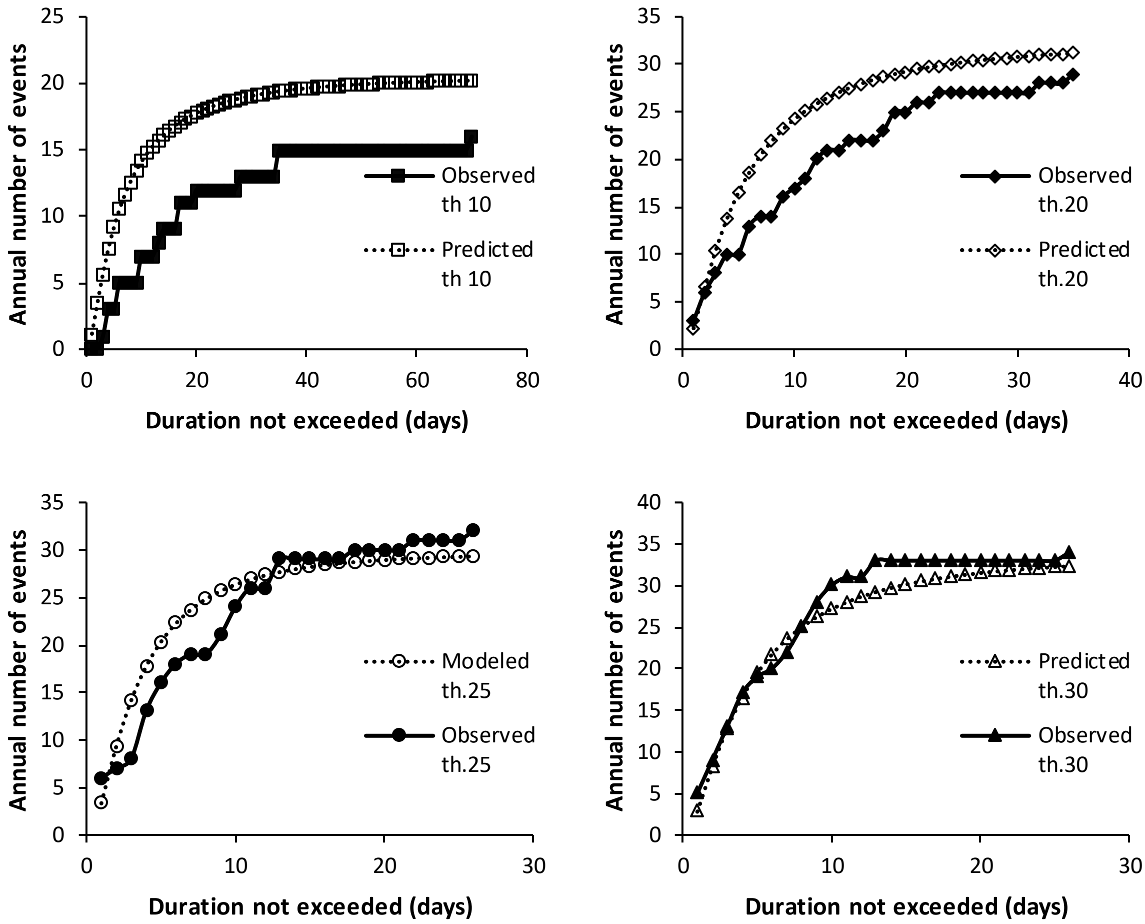


Fig. 28. Comparison between predicted by the eq. 12 and observed annual number of events of a given duration for the thresholds of 10, 20, 25 and 30 μg m^{-3}. Site of Centar, year 2012.

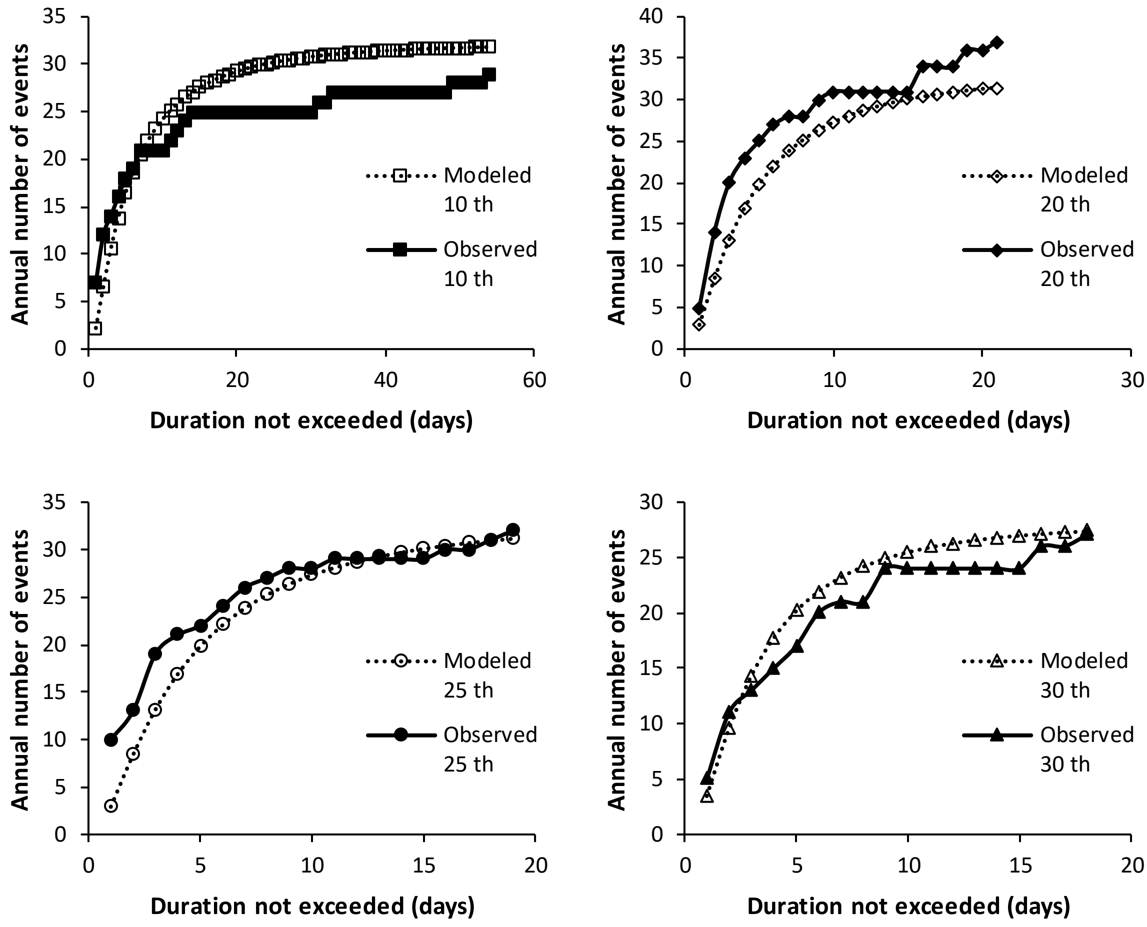


Fig. 29. Comparison between predicted by the eq. 12 and observed annual number of events of a given duration for the thresholds of 10, 20, 25 and 30 µg m^{-3}. Site of Milano Pascal, year 2015.

3.5. Models application case-study

Eqs. 3, 4, 5, 6, 7, 10, 11, 12 applied in this case-study were all established on a ratio (R_{TC}) between a concentration threshold T and the annual average PM2.5 concentration Ca, R_{TC}=T/Ca. Because of this characteristic one could calculate:

- the Total time of exceedance (TE_T) seen in eq. 3.
- the Number of Events (NE_T) as seen in eq. 4.
- the two parameters of the Lognormal distribution model (μ_T, σ_T) seen in eqs. 5 and 6.
- the lognormal distribution model ($F(D_T)$) providing the cumulative distribution of durations seen in eq. 7.
- the average duration (μ_{DT}) and the most frequent duration (M_{DT}) of the exceedance events seen in eqs. 10 and 11.
- the model for the number of events of a given duration ($NE_{D,T}$) seen in Eq. 12.

Furthermore, by using these formulations one could make predictions as to when is expected an event of high concentration of PM2.5. The predictions are useful when:

- there is a malfunction in one, or more of the measuring stations.
- one wants to reduce the annual average concentration of PM2.5 by enforcing measures throughout the year for the said reductions in order to reach a target annual average concentration that will be in compliance with the AQS limit concentration.

The second example represents what these equations are used for in this particular case-study where the data presented is from 2 different regions in Europe, Lombardy and Macedonia. The model incorporates the concentrations from the two regions and in the end by appointing a target annual average concentration the model computes the number of events and total time for that target annual average concentration. Because the AQ limit for the PM2.5 is on the annual average concentration and there is no limit for the number of exceedances the target limit is equal to 25 µg m^{-3} PM2.5. However, the European (Union, 2008), as well as Macedonian law for Air Quality state that the annual average concentration of PM2.5 should be 25 µg m^{-3} for the time being and further reduced to 20 µg m^{-3} by the 1st of January 2020. In this case-study these two annual average concentrations plus another one have been used as target for the PM2.5. The last annual average concentration target is 10 µg m^{-3}, because the World Health Organization (WHO, 2005), states that for this PM2.5 limit the adverse effects in the long run are reduced to minimum.
Once the target PM2.5 annual average concentration is determined one could start utilizing the eqs. 3, 4 and 12 in order to assess the Total time: TE_T, Number of Events: NE_T and the cumulative distribution of durations of those events: $NE_{D,T}$.

For the target of 25 μg m^{-3} PM2.5 the predicted and the observed cumulative distribution of the number of events and the total time for the thresholds can be seen on the Fig. 30.
When taking in account annual average concentration of 25 μg m^{-3} there are still some expected exceedances of all the 4 thresholds (10, 20, 25 and 30 μg m^{-3}).

Predictions for the 10 μg m^{-3} threshold are the following:

- total of 32 number of events
- average duration of 0.4 days
- maximum duration of 4.4 days
- total time of 260.7 days

Predictions for the 20 μg m^{-3} threshold are the following:

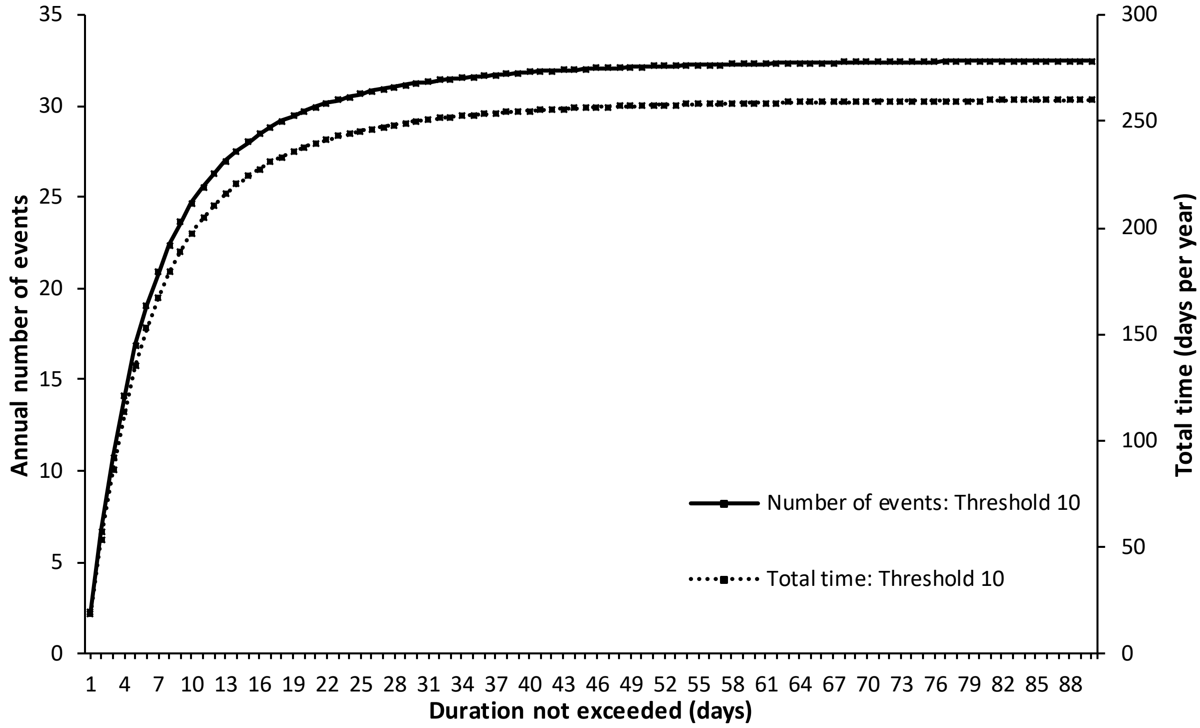
- total of 30 number of events
- average duration of 0.8 day
- maximum duration of 5.9 days
- total time of 153.8 days

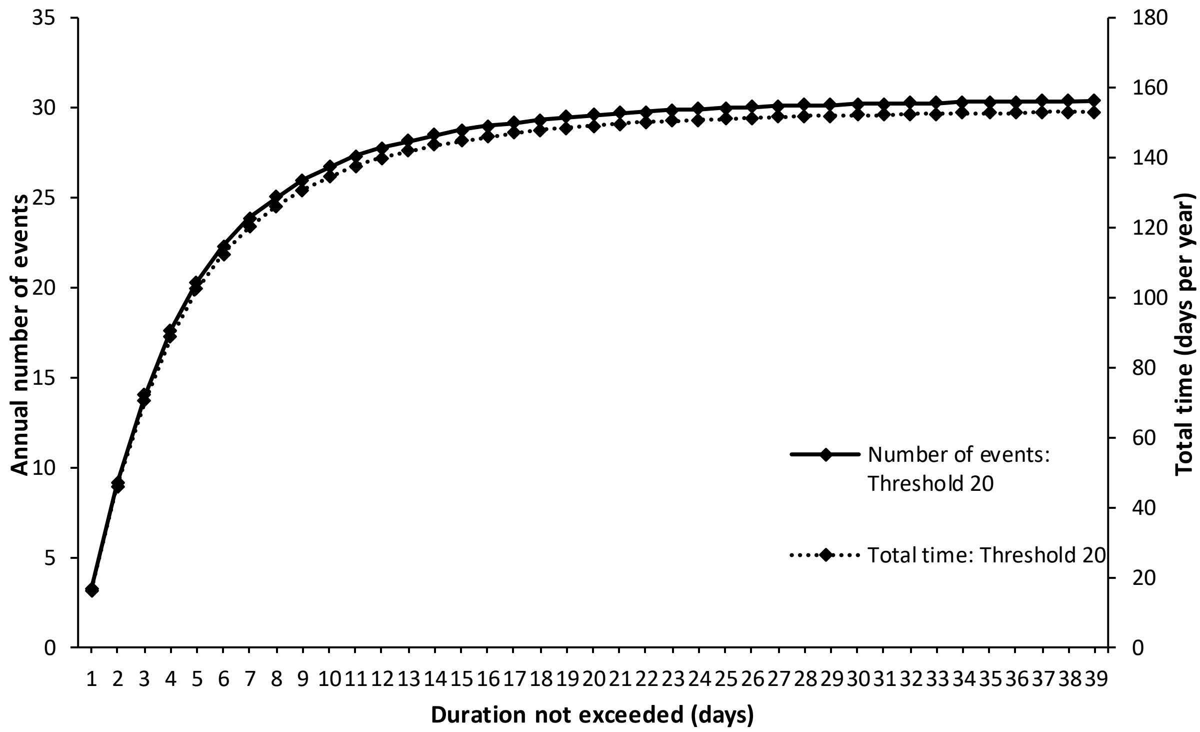
Predictions for the 25 μg m^{-3} threshold are the following:

- total of 27 number of events
- average duration of 0.6 day
- maximum duration of 6.2 days
- total time of 118.2 days

Predictions for the 30 μg m^{-3} threshold are the following:

- total of 23 number of events
- average duration of 1.2 day
- maximum duration of 6.3 days
- total time of 90.8 days


Annual number of events
Total time (days per year)
Duration not exceeded (days)
Number of events: Threshold 10
Total time: Threshold 10


Annual number of events
Total time (days per year)
Duration not exceeded (days)
Number of events: Threshold 20
Total time: Threshold 20

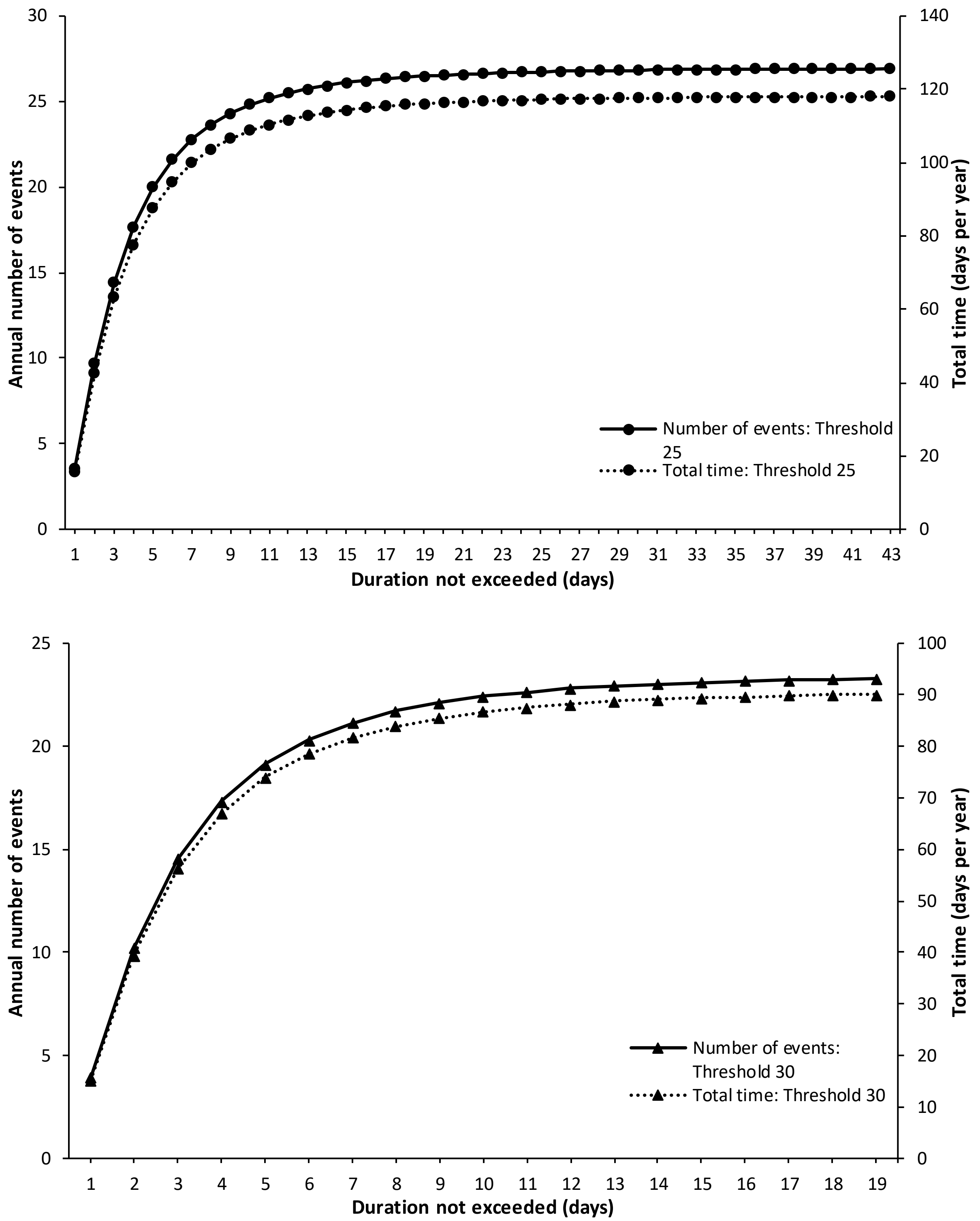


Fig. 30. Predicted cumulative distribution of the number of events and the total time for the thresholds of 10, 20, 25 and 30 μg m^{-3}. Target annual PM2.5 average concentration: 25 μg m^{-3}.

For the target of 20 µg m^{-3} PM2.5 required to be met by the 1st of January 2020, the predicted and the observed cumulative distribution of the number of events and the total time for the thresholds can be seen on the Fig. 31.
When taking in account annual average concentration of 20 µg m^{-3} there are still some expected exceedances of all the 4 thresholds (10, 20, 25 and 30 µg m^{-3}).

Predictions for the 10 µg m^{-3} threshold are the following:

- total of 33 number of events
- average duration of 1 day
- maximum duration of 5 days
- total time of 228.5 days

Predictions for the 20 µg m^{-3} threshold are the following:

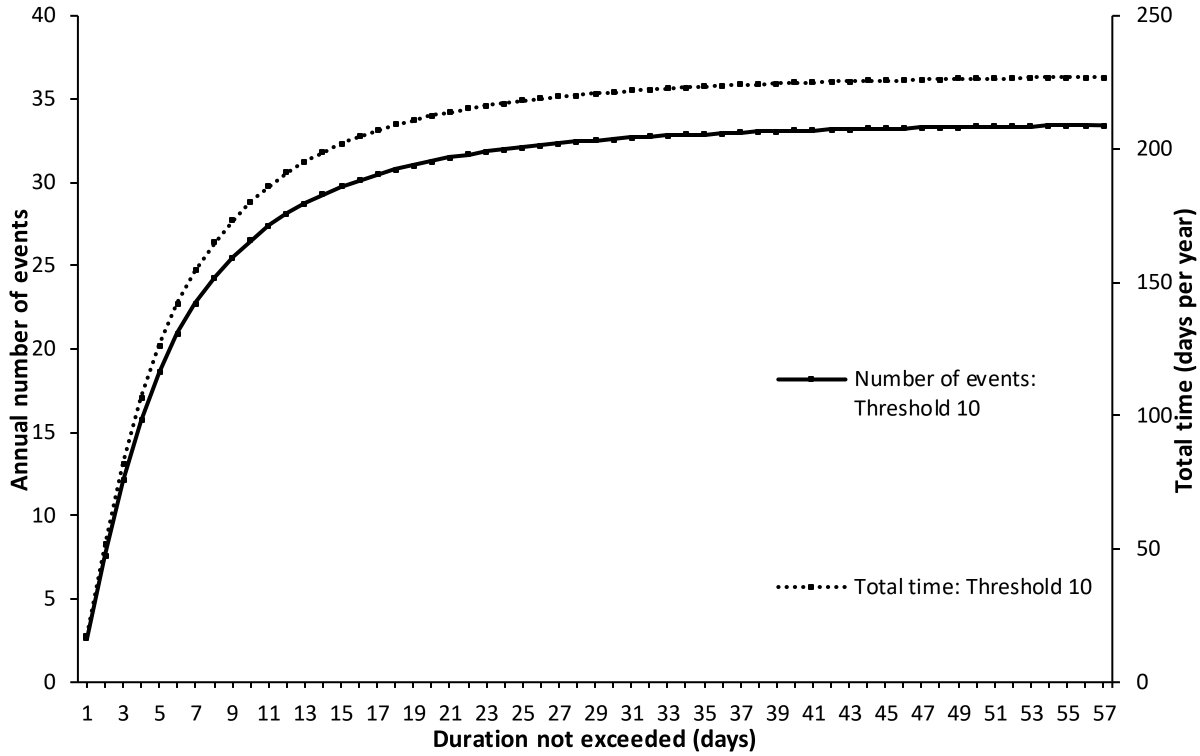
- total of 27 number of events
- average duration of 0.8 day
- maximum duration of 6.2 days
- total time of 118.2 days

Predictions for the 25 µg m^{-3} threshold are the following:

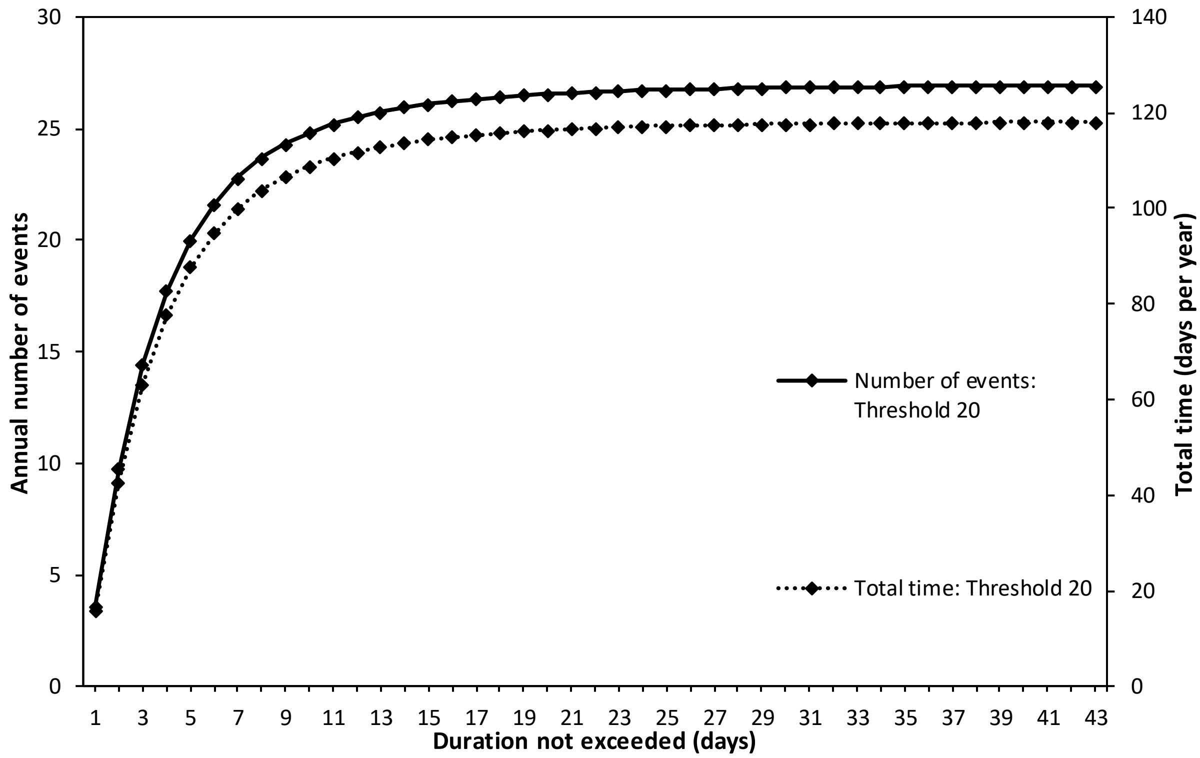
- total of 22 number of events
- average duration of 1.4 day
- maximum duration of 6.3 days
- total time of 85 days

Predictions for the 30 µg m^{-3} threshold are the following:

- total of 18 number of events
- average duration of 1 day
- maximum duration of 6.2 days
- total time of 61.1 days



Annual number of events
Total time (days per year)
Duration not exceeded (days)
Number of events: Threshold 10
Total time: Threshold 10



Annual number of events
Total time (days per year)
Duration not exceeded (days)
Number of events: Threshold 20
Total time: Threshold 20

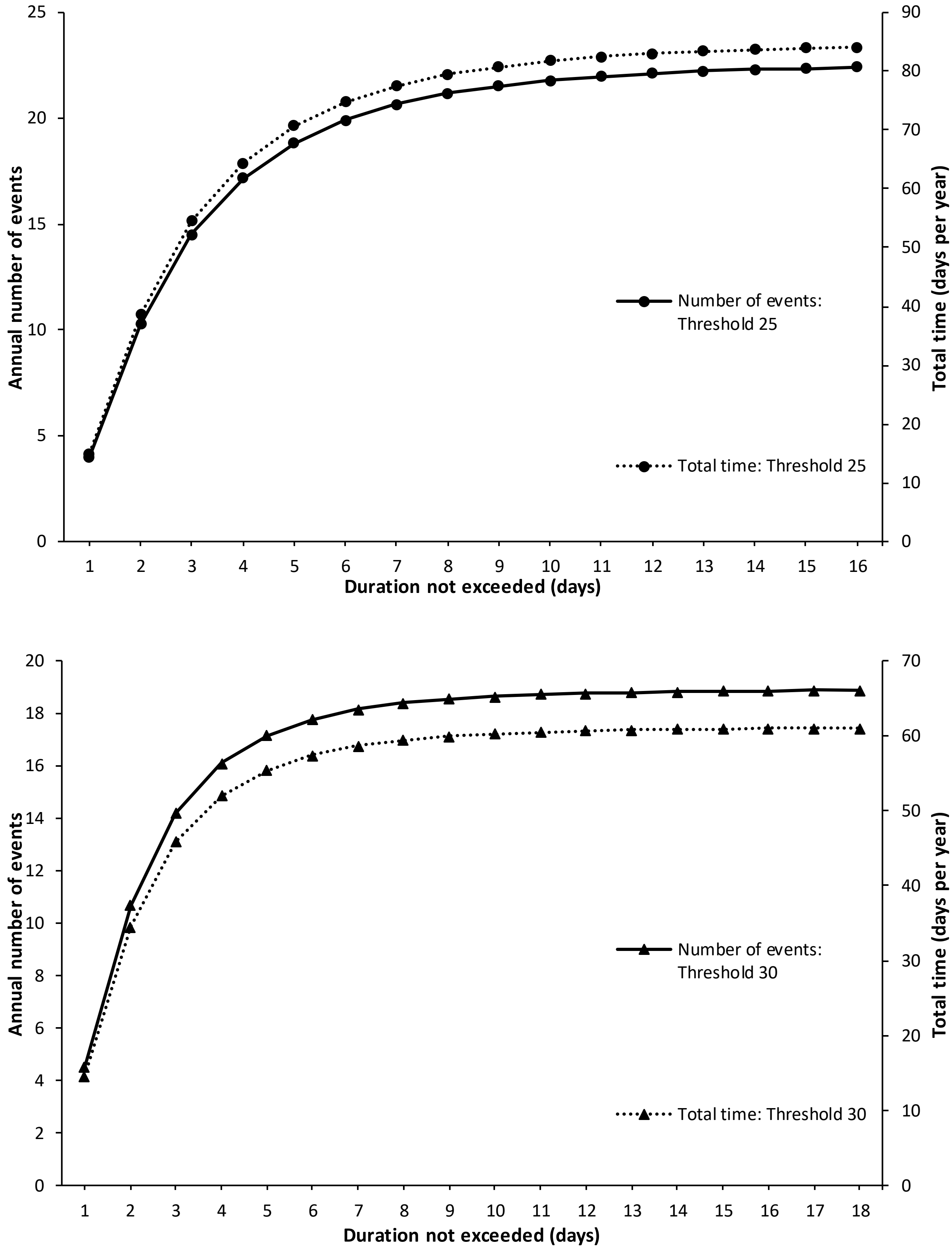


Fig. 31. Predicted cumulative distribution of the number of events and the total time for the thresholds of 10, 20, 25 and 30 µg m^{-3}. Target annual PM2.5 average concentration: 20 µg m^{-3}.

For the target of 10 µg m^{-3} PM2.5 stated by the World Health Organization (WHO, 2005), the predicted and the observed cumulative distribution of the number of events and the total time for the thresholds can be seen on the Fig. 32.
When taking in account annual average concentration of 20 µg m^{-3} there are still some expected exceedances of all the 3 thresholds (10, 20 and 25 µg m^{-3}), and for the 30 µg m^{-3} threshold there are no predicted events not time of exceedance.

Predictions for the 10 µg m^{-3} threshold are the following:

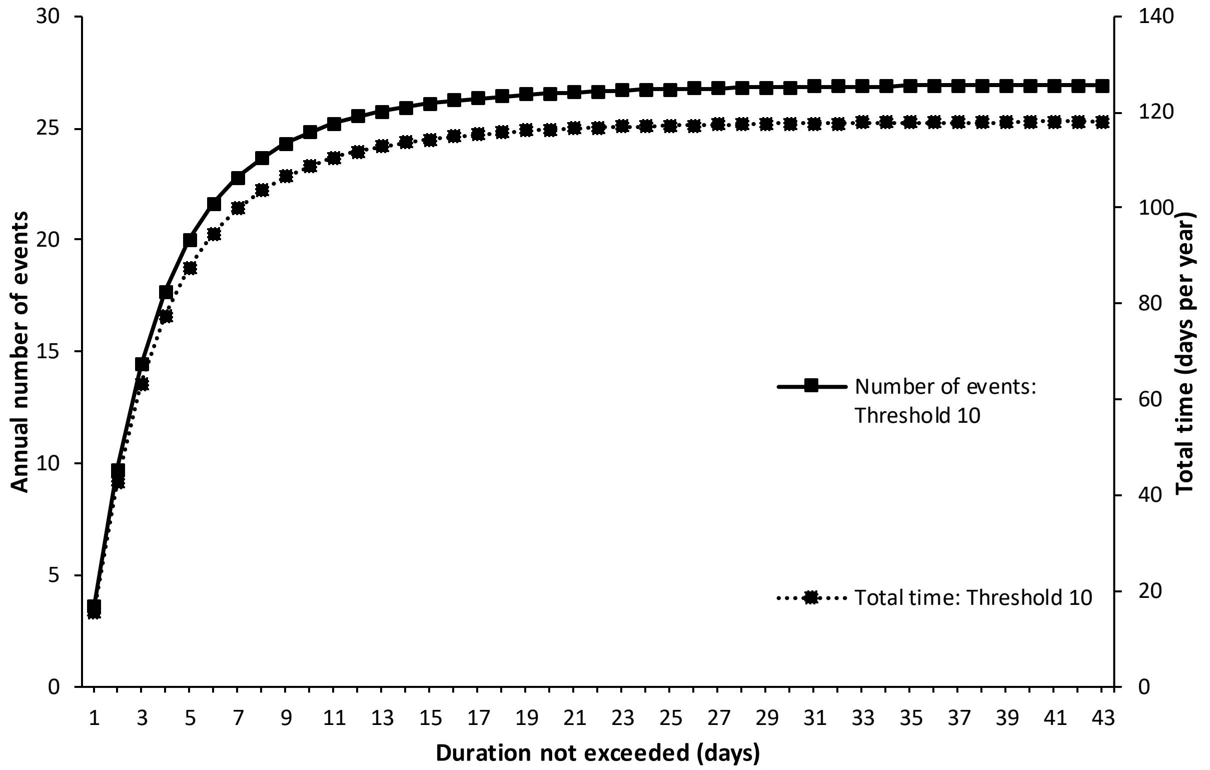
- total of 27 number of events
- average duration of 0.6 day
- maximum duration of 6.2 days
- total time of 118.2 days

Predictions for the 20 µg m^{-3} threshold are the following:

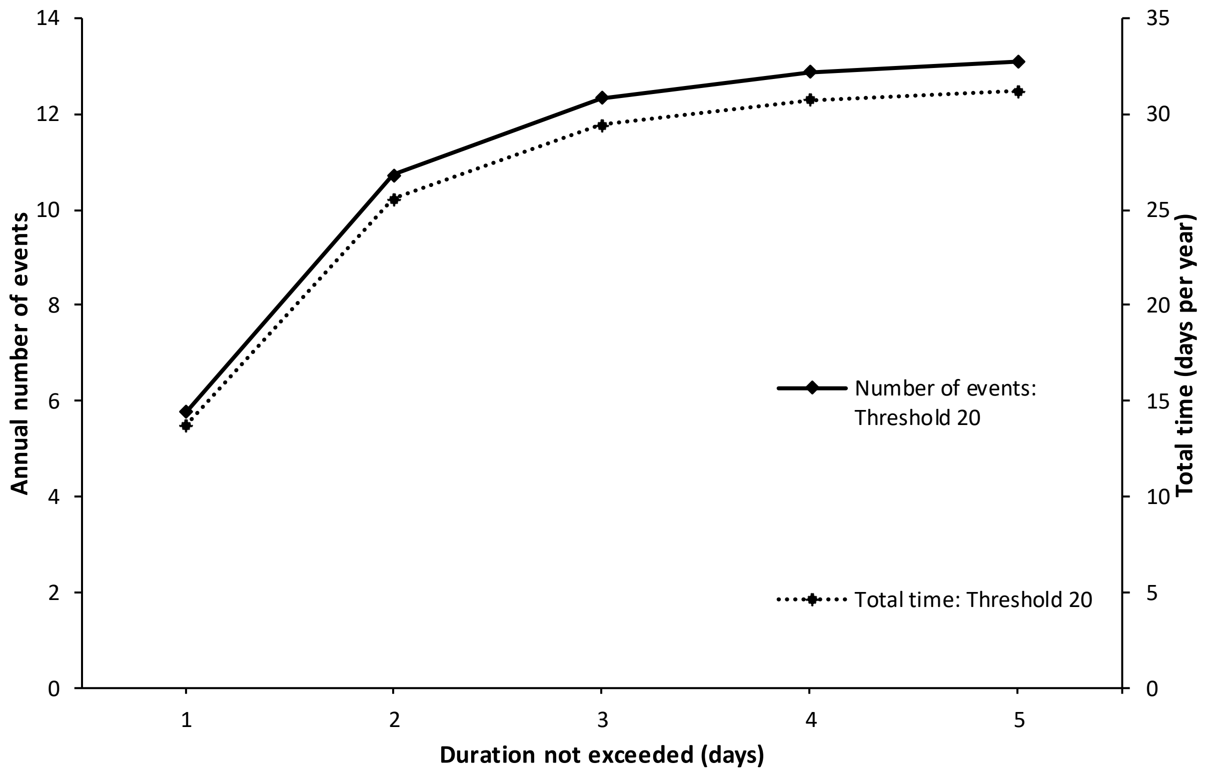
- total of 13 number of events
- average duration of 2.6 day
- maximum duration of 5.8 days
- total time of 31.6 days

Predictions for the 25 µg m^{-3} threshold are the following:

- total of 9 number of events
- average duration of 1.6 day
- maximum duration of 7 days
- total time of 16.3 days



30
25
20
15
10
5
0
140
120
100
80
60
40
20
0
1 3 5 7 9 11 13 15 17 19 21 23 25 27 29 31 33 35 37 39 41 43
Annual number of events
Total time (days per year)
Duration not exceeded (days)
Number of events: Threshold 10
Total time: Threshold 10



14
12
10
8
6
4
2
0
35
30
25
20
15
10
5
0
1 2 3 4 5
Annual number of events
Total time (days per year)
Duration not exceeded (days)
Number of events: Threshold 20
Total time: Threshold 20

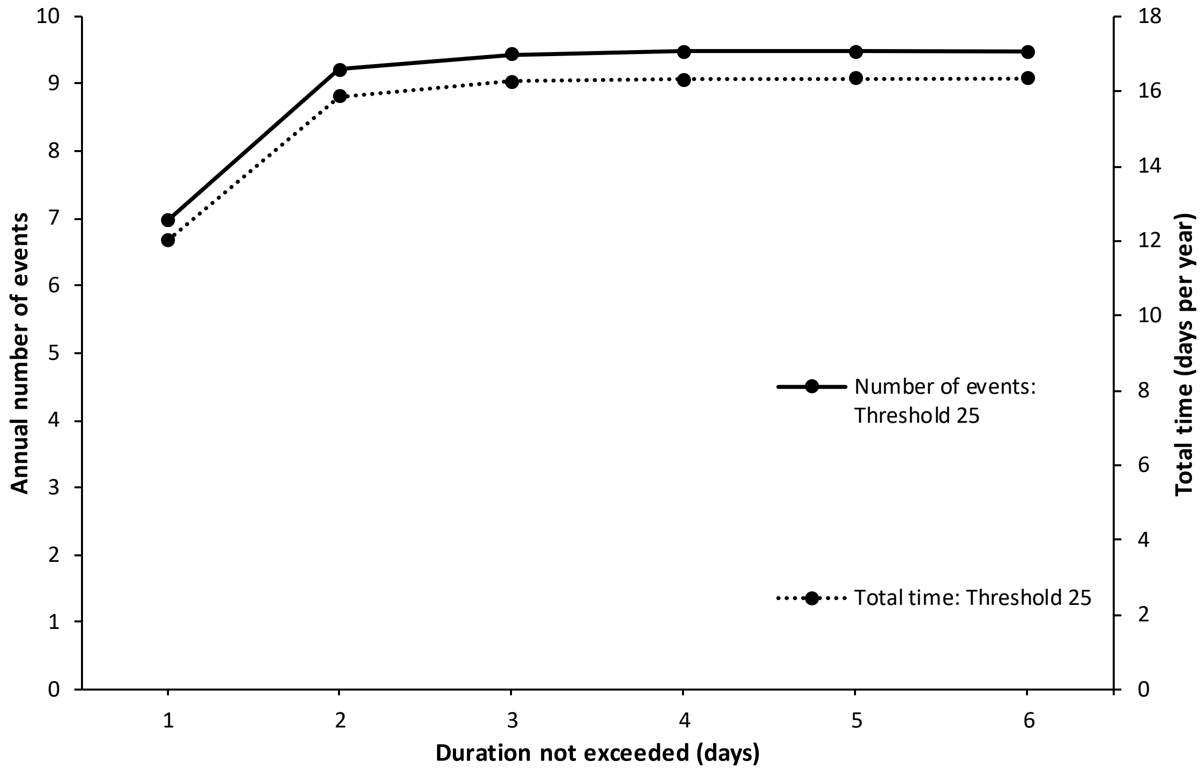


Fig. 32. Predicted cumulative distribution of the number of events and the total time for the thresholds of 10, 20, 25 and 30 µg m-3. Target annual PM2.5 average concentration: 10 µg m-3.

Conclusion

In this case study annual time series of daily average PM2.5 concentrations from measuring stations in two regions, were used, Lombardy and Macedonia. There were 11 measuring stations, of which 9 are in Lombardy and 2 in Macedonia. Each station measured daily average concentrations of PM2.5 from 2012 to 2017, in total 55 annual data sets. What was observed was that the daily limit concentration as well as the annual limit concentration of PM2.5 ,25 µg m^{-3}, was continuously exceeded for a certain period. These exceedances of the daily and annual PM2.5 concentration limit signify that the annual time series will not be in compliance with that limit.

In order to assess how much certain thresholds are exceeded and for how long, the following was calculated:

- Total time of exceedance – or the time in which the concentration of PM2.5 is above a certain threshold concentration,
- Number of events – or how many times the concentration of PM2.5 has exceeded a certain threshold,
- and the distribution of the durations.

The TE_T and NE_T presented here in this case-study rely on R_{TC} - ratio of the threshold T and the annual average concentration Ca. If one goes back to eqs. 3 and 4 one would notice that the entire formulation is dependent on the R_{TC}. As a result of the said formulations there are the models depicting the observed data. For the distribution of the durations exceeding a threshold however the two-parameter lognormal model was used because almost all the natural phenomena follow this distribution. The accuracy of the two-parameter lognormal model is calculated by the Fractional Bias and the Fractional Error as well as the Correlation coefficient between the predicted and observed data, eqs. 1 and 2. The two parameters of the lognormal model were used in a linear regression as a function of the R_{TC}, in order to develop the cumulative distribution model of durations established on the annual concentration Ca.

The ratio between the threshold concentration and the annual average concentration in the end was used to estimate episode events and their duration of high concentration levels of PM2.5, when one needs to see outcomes of different scenarios of expected annual average concentration of PM2.5.

This case-study was designed for Lombardy and Macedonia because the data used were from these sites. With the combination of the developed models for the cumulative distribution of durations and the number of events exceeding a threshold, a prediction of the cumulative annual number of events exceeding a given concentration threshold for a time lower than or equal to the duration was done for different target annual average concentrations: 25, 20 and 10 µg m^{-3}. Depicting 3 scenarios:

- 25 µg m^{-3} is the current annual average limit concentration.

- 20 μg m^{-3} is the near future annual average limit concentration that should be met by the 1st of January 2020.
- 10 μg m^{-3} is an annual average concentration suggested by the World Health Organization for the lowest adverse effects of the pollution.

For these 3 target annual average concentrations the model predicts exceedances of all the 4 thresholds (10, 20, 25 and 30 μg m^{-3}), except for the lowest annual average concentration target (10 μg m^{-3}) where there are no exceedances of the 30 μg m^{-3} threshold.

References

Caserini, S., 2004. *Primary PM10 and PM2.5 Emission Inventory in Lombardy,* s.l.: s.n.

EPA, 2014. *"Positive Matrix Factorization (PMF) 5.0. Fundamentals and User Guide".* Washington, DC 20460: U.S. Environmental Protection Agency.

G. Lonati, S. C. M. G., 2011. The duration of PM10 concentration in a large metropolitan area. *Atmospheric Environment,* Issue 45, pp. 137-146.

Georgopoulos, G. S. H., 1982. Statistical distributions of air pollutant concentrations. *Environmental Science and Technology ,* Issue 16, pp. 401A-416A.

Giugliano, M. C. S. M. F., 1998. The duration of high NO2 and CO concentrations in an urban atmosphere. *Atmospheric Environment ,* Issue 32, pp. 2923-2929.

Kelly, F. J. &. F. J. C., 2012. *Size, source and chemical composition as determinants of toxicity attributable to ambient particulate matter. ATMOSPHERIC ENVIRONMENT, 60(0), 504 - 526.,* s.l.: s.n.

Lancet, T., 2016. Global, regional, and national comparative risk assessment of 79 behavioral, environmental and occupational, and metabolic risks or clusters of risks, 1990–2015: a systematic analysis for the Global Burden of Disease Study 2015.. 7 October, p. 388:1659–1724..

Pope CA 3rd, E. M. D. D., 2009. Fine particulate air pollution and life expectancy in the United States. *New England Journal of Medicine,* pp. 360(4):376-386.

Ryan, P., 1991. An overview of human exposure modeling. *Journal of Exposure Analysis and Environmental Epidemiology ,* Issue 1, pp. 453-474.

Ryan, P. S. J. L. R., 1986. Estimating personal exposure to NO2. *Environment International,* Issue 12, pp. 395-400.

S. Koton, N. M. Y. V. M. D. M. B. Y. D. D. M. S. Y. G., 2013. Cumulative Exposure to Particulate Matter Air Pollution and Long-Term Post-Myocardial Infarction Outcomes. *Preventive Medicine 57,* pp. 339-344.

Seinfeld, J. & Pandis, S., 1998. *Atmospheric Chemistry and Physics: From Air Pollution to Climate Change (2nd ed.)..* Hoboken, New Jersey: John Wiley & Sons, Inc. p. 97. ISBN 0-471-17816-0..

Stephanou, E., 2012. Aerosols PM10 and PM2.5. Volume 1, pp. 191-199.

Sun, Y. Q. J. Z. W. P. F. J. L. T. Y. a. Y. Y., 2014. Investigation of the sources and evolution processes of severe haze pollution in Beijing in January 2013. *Journal of Geophysical Research: Atmospheres,* 119(7), p. 4380–4398.

Union, E., 2008. Directive 2008/50/EC of the European Parliament and of the Council of 21 May 2008 on ambient air quality and cleaner air for Europe. *Official Journal of the European Union.*

Union, E., 2013. *European Union emission inventory report 1990–2014 under the UNECE Convention on Long-range Transboundary Air Pollution (LRTAP),* s.l.: s.n.

Websites accessed

EPA., 2014. *EPA.* [Online]
Available at: https://www.epa.gov/air-trends/particulate-matter-pm10-trends

EPA., 2017. *EPA.* [Online]
Available at: https://www.epa.gov/air-research/history-air-pollution, accessed

OSHA., 2017. *OSHA.* [Online]
Available at: www.osha.gov
[Accessed March 2017].

ARPA, 2017. *ARPA.* [Online]
Available at: http://www2.arpalombardia.it/sites/qaria/_layouts/15/qaria/RicercaIDati.aspx
[Accessed March 2017].

MOEPP, 2014. *MOEPP.* [Online]
Available at: http://www.moepp.gov.mk/wp-content/uploads/2014/10/UREDBA-ZA-IZMENUVAN%D0%88E-IDOPOLNUVANJE-NA-UREDBATA-ZA-GRANICNIVREDNOSTI.pdf
[Accessed March 2017].

MOEPP, 2015. *MOEPP, , accessed.* [Online]
Available at: http://www.moepp.gov.mk/wp-content/uploads/2014/11/Statistika-na-Zivotna-Sredina-2015.pdf
[Accessed March 2017].

MOEPP, 2017. *MOEPP.* [Online]
Available at: http://airquality.moepp.gov.mk/?page_id=469
[Accessed March 2017].

WHO, 2005. *Air Quality Guidelines, Global Update 2005.* [Online]
Available at:
http://apps.who.int/iris/bitstream/10665/69477/1/WHO_SDE_PHE_OEH_06.02_eng.pdf
[Accessed 2017].