UNIVERSITÀ DEGLI STUDI DI BERGAMO
DIPARTIMENTO DI INGEGNERIA DELL'INFORMAZIONE
E METODI MATEMATICI°

QUADERNI DEL DIPARTIMENTO

Department of Information Technology and Mathematical Methods

Working Paper

Series "*Mathematics and Statistics*"

n. 4/MS – 2007

Comparing air quality among Italy, Germany and Poland using BC indexes

by

Olha Bodnar, Michela Cameletti, Alessandro Fassò, and Wolfgang Schmid

° Viale Marconi. 5, I – 24044 Dalmine (BG), ITALY, Tel. +39-035-2052339; Fax. +39-035-562779

Comparing air quality among Italy, Germany and Poland using BC indexes

Olha Bodnar[1], Michela Cameletti[2], Alessandro Fassò[2], Wolfgang Schmid[1]
[1]Europa Universität Viadrina - Frankfurt (Oder), [2]University of Bergamo

Abstract

In this paper we discuss air quality assessment in three Italian, German and Polish regions using the index methodology proposed in Bruno and Cocchi (2002, 2007). This analysis focuses first on the local air quality situation of each considered country and then adopts a more general approach with a comparing purpose in terms of pollution severity and toxicity. This is interesting in a global European perspective where all countries are commonly involved in assessing air quality and taking proper measures for improving it. In this context, air quality indexes result to be a powerful data-driven tool which are easily calculated and summarize a complex phenomenon, such as air pollution, in promptly understandable indicators. In particular, the main objective of this work is to evaluate the index performances in discriminating different air pollution patterns. This kind of analysis can be particularly useful, for example, in the perspective of constructing an indicator of air pollution.

1 Introduction

Air quality is known to be an important issue for both the governments and the citizens. On the one hand, the latter are interested in detailed and timely air quality information regarding their own territory. For example, the European Environment Agency (2007) gives detailed behavioral hints for high tropospheric ozone events. On the other hand, for example, the EU member states have to comply with European and national directives which fix limit values and alert thresholds for the main pollutants and provide criteria and reference methods for measuring most of the relevant air pollutants. On this subject, see the EU Council Directive 1999/30/EC, relating to pollutant limit values, and EU Council Directive 1996/62/EC, on ambient air quality assessment and management.

In this context a simple and effective tool, such as air quality indexes, is needed for giving timely and easy to communicate information about air quality, assessing compliance with reference standards and evaluating the effects of emissions control policies. As a matter of fact, air quality indexes are easily computed and synthesize multiple and multiscale measurements in a standardized indicator that provides timely and easily understandable information. Their use is suggested, for example, by the U.S. Environmental Protection Agency (EPA) which publishes national guidelines for their computing and reporting (U.S. EPA, 2003).

Although the European directives define the measurement methods for various pollutants, there are differences among the national monitoring networks in terms of spatial distribution of the various instruments and, hence, of the monitored pollutants. Moreover, especially with long time series and trend analysis, network characteristics change both in space and time, thus giving rise to heterogeneous networks (Fassò *et al.*, 2007).

From the scientific investigation point of view, air quality indexes can be used for preliminarily as a first step to more complex, possibly model-based, analysis that aim, for example, at air quality spatio-temporal modeling and mapping or to evaluate the impact of air pollution exposure on human health (Bellini *et al.*, 2007; Englert, 2004; Pope, 2000). Moreover, they can be used as sub-indicators in composite indicators, see e.g. Saisana *et al.* (2005) and references therein. Recently Lagona (2005) and Chiu *et al.* (2007) have proposed an approach to indexes by means of the latent factors of a Hidden Markov Model. Although it is a promising approach, simplicity and interpretability are still under study and we opt here for explicit index definition.

In this work, we use the BC indexes methodology proposed by Bruno and Cocchi (2002, 2007) for assessing and comparing air quality in three regions from Italy, Germany and Poland. These countries are known to have different geo-meteorological characteristics and different population densities giving also markedly different pollution levels. Hence, it is interesting to understand at which extent a BC index can point out seasonality and discriminate among different air pollution patterns. In particular, the case of heterogeneous monitoring networks is discussed with reference to the BC indexes showing which one is preferable for comparing perspectives.
The structure of the work is the following: in Section 2, we present the considered Italian, German and Polish regions together with some relevant geographical and anthropic characteristics. Moreover, the monitoring networks used in year 2005 are discussed in terms of the spatial distribution

of stations and pollutants sensors. In Section 3, we introduce the notation and methodology of BC indexes together with some comments about their interpretation. The results are given in Section 4, where the obtained index time series are widely discussed within and between the considered areas. In particular, focusing on the index performance in terms of the capacity of discriminating different air pollution situations, we show how the indexes are related to the monitoring network structure.

2 Data description

The index analysis, referring to year 2005, is performed on the Piedmont and Lombardy regions in Italy, on the Berlin and Brandenburg states in Germany and on the Masovian Province in Poland, which are discussed in the following subsections.

Following the above mentioned European directives, we consider the pollutants listed in Table 1 together with the corresponding standard limit values and the temporal aggregation functions used for the indexes of Section 3.

The pollutant considered are related with industrial, domestic and traffic sources. In particular, sulphur dioxide (SO_2) is an "old pollutant"as it is mainly the result of coal burning which has been replaced in most European countries; nevertheless it is still monitored because its high potential impact on both humans and the environment.

Nitrogen dioxide (NO_2), Carbon monoxide (CO) and benzene are strongly related with combustion, road traffic and petrol distribution.

Particulate matters with an aerodynamic diameter lower than 10 μm (PM_{10}) do not include relevant pollution measurements, such as ultrafine particulate matters ($PM_{2.5}$ and PM_1) which have been proved to be health risk factors, because of scarce availability for the considered year.

Last but not the least, tropospheric ozone (O_3) is a secondary pollutant produced by reaction between possibly transported nitrogen dioxide, hydrocarbons and sunlight. It is known to be especially high on sunny hills and mountains around highly trafficated areas, as in the Italian case, and has a very skewed distribution and a complex dynamics, see e.g. Fassò and Negri (2002).

2.1 Italian region

We consider Piedmont and Lombardy regions, covering 49.260 km^2 in the western part of the so called Po Valley, North of Italy, as shown in Figure 1. The considered area stretches for about 300 km in the east-west direction

Pollutant	Measurement unit	Temporal aggregation function	Standard limit
Benzene	mg/m^3	Daily average	10 mg/m^3
CO	$\mu g/m^3$	Daily max of 8-hours moving averages	10 $\mu g/m^3$
NO_2	$\mu g/m^3$	Daily maximum	300 $\mu g/m^3$
PM_{10}	$\mu g/m^3$	Daily average	50 $\mu g/m^3$
O_3	$\mu g/m^3$	Daily max of 8-hours moving averages	120 $\mu g/m^3$
SO_2	$\mu g/m^3$	Daily average	125 $\mu g/m^3$

Table 1: Information about considered pollutants.

and is surrounded by the Alps on the northern and western sides, by the Apennines on South and by plain on East. Note that the mountain chains form a sort of c-shaped barrier that protects the area from the major air circulation. For this reason, especially during winter, air tends to stagnate and this leads to pollutant accumulation and observed high air pollution concentration levels. Moreover, the Po Valley is characterized by the presence of large and densely populated urban centers and metropolitan areas with a busy motorways network. The anthropic impact can be related to the density of the population which amounts to 284 persons per km^2 and grows up to 486 persons per km^2 if we exclude mountain areas.

The monitoring networks of both regions are managed by the corresponding regional environmental agencies which are responsible for the air quality monitoring, the public information and data supply. As shown in Table 2, there are 127 monitoring stations in Lombardy and 72 in Piedmont. More than 90% of the stations is of urban type, which means that they are located in commercial and residential zones characterized by high traffic levels.

The network spatial distribution is related more to human risk than pure spatial coverage. As a result, stations are mainly located in the highly populated provinces of the two chief towns, that is Milan, with 33% of Lombardy stations, and Turin, with 42% of Piedmont stations.

Nevertheless, as it can be seen in Figure 1, the network spatial coverage is good and stations can be found also in plain rural areas and urbanized alpine valleys. Despite of this, considering the monitored pollutants, Table 3 shows that some are intensively monitored, namely CO and NO_2, which are considered for local acute events, while others less, namely O_3 and PM_{10}, which are sampled mainly on a spatial representative basis, and last benzene which is scarcely monitored, especially in Lombardy. We term unbalanced such an heterogeneous network.

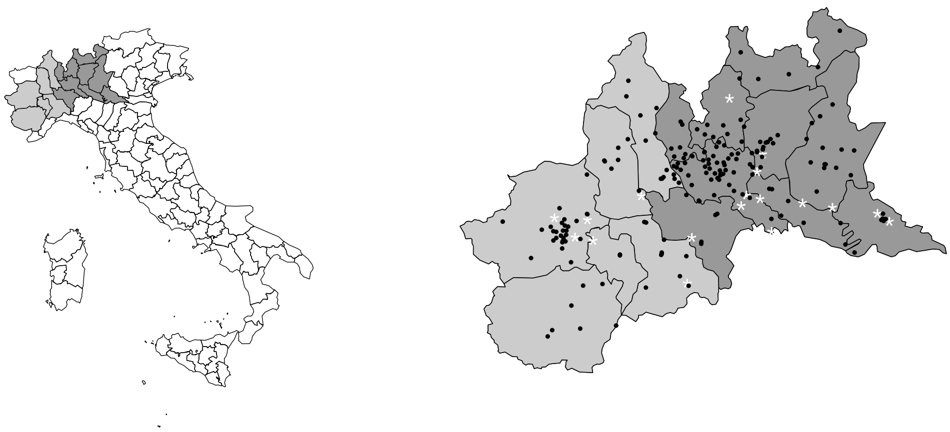


Figure 1: Left side: Piedmont (western, light gray) and Lombardy (eastern, dark gray) location. Right side: pollutant monitoring network (white stars for rural stations and black points for urban ones).

Type of Station	Piedmont	Lombardy	Total
Rural	6	12	18
Urban	66	115	181
Total	72	127	199

Table 2: Piedmont and Lombardy monitoring network description according to the station type.

Pollutant	Piedmont	Lombardy	Total
Benzene	14	7	21
CO	43	81	124
NO_2	63	121	184
PM_{10}	33	46	79
O_3	29	57	86
SO_2	28	47	75
Total	210	359	569

Table 3: Piedmont and Lombardy pollutant sensors .

2.2 German region

We consider the Berlin-Brandenburg region which is located in the eastern part of Germany and consists of the Brandenburg federal state and the national capital Berlin. Its total extension is 30.370 km^2 and the population density is 195.8 persons per km^2. If we consider only the Berlin metropolitan area, whose extension is 891 km^2, the population density is 3821 persons per km^2.

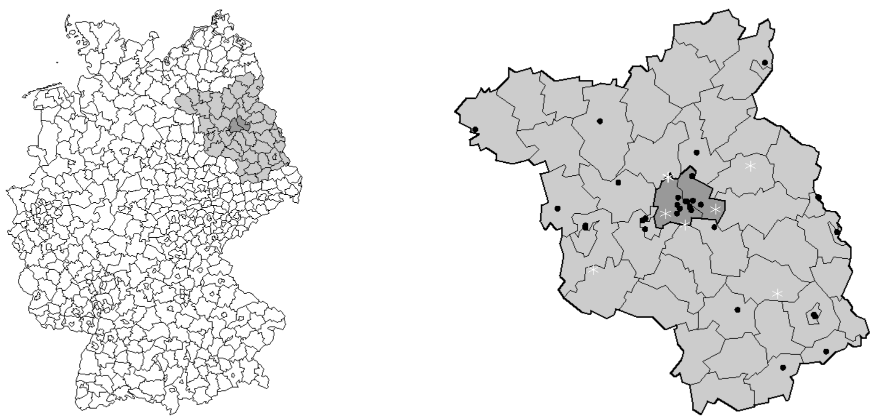


Figure 2: Left side: Berlin (dark gray) and Brandenburg (light gray) location. Right side: pollutant monitoring network (white stars for rural stations and black points for urban ones).

The Berlin-Brandenburg region is located in the North European Lowlands which slope North toward the Baltic Sea, with the northern lowlands being very flat, below sea-level in parts. Most of the Berlin-Brandenburg region lies well under 100 meters above sea level with hills hardly reaching 200 meters. South, the Central German Uplands rise a relevant height only faraway and has no air circulation reduction effect for the Berlin-Brandenburg region.

Eastward the exterminated North and East European Lowlands do not prevent Atlantic air circulation and the weather in the Berlin-Brandenburg region is not stable and predictable. Low and high pressure systems change quickly. According to the Köppen-Geiger climate classification (Peel *et al.*, 2007) Berlin and Brandenburg has a temperate/mesothermal climate (Cfb). The climate is influenced by dry continental air masses from Eastern Europe and by maritime air masses from the Atlantic. For these reasons the pollutants do not accumulate in atmosphere for long periods and the concentrations are usually moderately low.

For instance the mean annual temperature in 2005 for Berlin is 9.4 °C (48.9 °F) and its mean annual precipitation totals 578 millimeters (26.8 in). The warmest months are June, July, and August, with mean temperatures of 16.7 to 17.9 °C (62.1 to 64.2 °F). The coldest are December, January, and February, with mean temperatures of -0.4 to 1.2 °C (31.3 to 34.2 °F).

As in the case of Italy, both the Berlin and Brandenburg monitoring networks are managed by the respective regional environmental agencies. There are 41 monitoring stations in the overall region with 18 (43.9%) allocated in Berlin and 23 (56.1%) in the Brandenburg federal state. In the case of Brandenburg, there are 3 rural stations, while for the Berlin case 33.3% of the stations are of the rural type (see Table 4). According to Table 5, both networks have a relatively homogeneous pollution coverage except for benzene which is scarcely monitored.

Type of Station	Berlin	Brandenburg	Total
Rural	6	3	9
Urban	12	20	32
Total	18	23	41

Table 4: Berlin and Brandenburg monitoring network description according to the station type.

2.3 Polish region

For the comparison study, we selected the central-eastern region of Poland named Masovian Province, where the capital Warsaw is located. It is the

Pollutant	Berlin	Brandenburg	Total
Benzene	4	2	6
CO	9	11	20
NO_2	14	22	36
PM_{10}	10	21	31
O_3	7	19	26
SO_2	7	13	20
Total	51	88	139

Table 5: Berlin and Brandenburg pollutant sensors.

largest and most populous province of Poland and occupies 35.598 km^2 with the total population density amounting to 144.3 persons per km^2. The Masovian region lies on the eastern part of the North European Lowlands and is covered by several large forest complexes with a temperate continental climate. The Köppen-Geiger classification is equal to Dfb. This means that in comparison to Berlin and Brandenburg the winters are colder and longer. In summer the temperatures are nearly the same, however, it is more rainy in the Masovian region. The mean temperature in the year 2005, for instance, in Warsaw is -2 °C (28 °F) in January and 18 °C (64 °F) in July. The annual rainfall averages 680 millimeters (26.8 in), the most rainy month is being July.

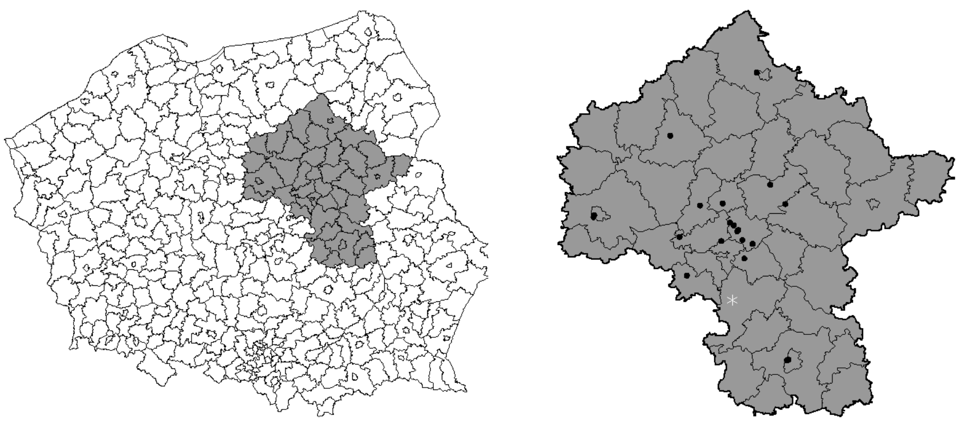


Figure 3: Left side: Masovian location. Right side: pollutant monitoring network (white stars for rural stations and black points for urban ones).

Type of Station	Masovian
Rural	1
Urban	20
Total	21

Table 6: Masovian monitoring network description according to the station type.

Pollutant	Masovian
Benzene	2
CO	7
NO_2	12
PM_{10}	18
O_3	8
SO_2	11
Total	58

Table 7: Masovian pollutant sensors.

Air pollution concentrations are examined in agreement with the Regulation of the Minister of Environment from June 2002 (Government Regulations and Laws Gazette n 87, item 798). According to the regulations concerning the national monitoring of the environment (Government Regulations and Laws Gazette n. 112, item 982), measuring data from different measuring stations and networks can be used for monitoring air quality. Of the 67 measuring stations working for the air monitoring network 21 had enough data for year 2005. Table 6 shows that 20 stations are of the urban type, while there is only one station of the rural type. Moreover, 6 (29%) stations out of 21 stations are situated in Warsaw. With regard to the monitored pollutants, Table 7 shows the spatial distribution, which is between the Italian and German ones, and has an high percentage of stations with PM_{10} sensors, a low percentage of O_3 sensors and the same sensor scarcity for benzene.

3 BC indexes methodology

Air quality data are defined over three dimensions regarding, respectively, the temporal (when?), the spatial (where?) and the pollutant (what?) definitions. In order to obtain daily air quality index time series, we aggregate the elementary data over the three dimensions. As described in Bruno and Cocchi (2002), after obtaining daily data by means of a temporal synthesis, it is possible to choose the order for the subsequent aggregations according to the purposes of the analysis. As our main objective is to compare air quality between the three regions with reference to health risk, we start by aggregating first among pollutants, taking the maximum among the standardized pollutants, and then we aggregate among stations. For aggregating data which refer to different pollutants we use the natural standardization procedure given by equation (2) of Bruno and Cocchi (2007), which is based on the standard limit values of Table 1.

To see this, let X_{spdh} be the elementary measurement which corresponds to the concentration of pollutant $p = 1, \ldots, P$, station $s = 1, \ldots, S$, day $d = 1, \ldots, D$ and hour $h = 1, \ldots, 24$. Note that it is not required that each pollutant is measured in all the S considered stations and that missing values are allowed. The first step is the temporal aggregation that is transforming hourly data into daily data X_{spd}; this is done using the temporal aggregating functions reported in the third column of Table 1 according to EU Directives.

Then, daily data X_{spd} are aggregated first by pollutant and then by station using the median (m) or the maximum (M) as aggregating function. In particular, the following BC indexes are computed for each day d

$$I\left(SP.MM\right) = I_d\left(SP.MM\right) = \max_s \left[\max_p \left(\frac{X_{spd}}{u_p}\right)\right] \tag{1}$$

$$I\left(SP.mM\right) = I_d\left(SP.mM\right) = \operatorname*{median}_s \left[\max_p \left(\frac{X_{spd}}{u_p}\right)\right] \tag{2}$$

where SP refers to the pollutant-station order of aggregation and u_p is the standard limit value of Table 1. Thanks to this, the indexes are defined on an a-dimensional scale where the unit is the reference value: indexes greater than one correspond to dangerous situations with exceedings of toxic stuff in atmosphere; obviously the higher the index value, the greater the level of air pollution and the greater the health concern.
Index (1), which is named the Maxmax index in the sequel, is given by the maximum value over stations of the maximum concentrations over pollutants and makes it possible to determine, for each day, the station corresponding to the maximum. This can be particularly useful for characterizing critical stations. On the other hand, index (2) is given by the median among stations of the maximum pollutant concentrations. It follows that the comparison between index $I\left(SP.MM\right)$ and $I\left(SP.mM\right)$ can be used for assessing the spatial or network variability. As a matter of fact, if index (2) is near index (1) the spatial median is near to the spatial maximum, which means spatial

homogeneity, and a severe air quality situation is to be referred to the whole considered area. On the other hand, if $I\,(SP.MM)$ differs markedly from $I\,(SP.mM)$ then spatial variability is high and the worst situation is related only to a reduced fraction of stations. So the one's complement ratio of the two indexes can be used for computing the following dispersion index

$$V = V_d = 1 - \frac{I_d\,(SP.mM)}{I_d\,(SP.MM)} \tag{3}$$

which is low in case of spatial or network homogeneity and increases when the spatial or network variability is higher reaching its maximum, one, when the median is equal to the maximum.

Other indexes, besides (1) and (2), can be promptly and easily calculated using quantiles different from the median, for example, the third quartile or the 90^{th} percentile. Moreover, an useful alternative to indexes (1) and (2) arises from using the station-pollutant aggregation order which leads to the following indexes:

$$\begin{aligned} I\,(PS.MM) &= I_d\,(PS.MM) = \max_p \left[\max_s \left(\frac{X_{spd}}{u_p} \right) \right] \\ I\,(PS.Mm) &= I_d\,(PS.Mm) = \max_p \left[\operatorname*{median}_s \left(\frac{X_{spd}}{u_p} \right) \right] \end{aligned} \tag{4}$$

Note that the Maxmax index is invariant with respect to the aggregation order, so that $I\,(SP.MM) = I\,(PS.MM)$, and can be considered as a benchmark because it corresponds to the worst air quality situation with respect to both space and pollutant. Hence, it is possible to use the Maxmax index for identifying the most severe pollutant for each day, which is also termed the decisive pollutant by Bruno and Cocchi (2002, 2007). These information can eventually be used by the governments in order to highlight which are the most dangerous pollutants and consequently to propose solution and programs that should be taken in order to reduce their emissions.

4 Discussion of the results

In this section we discuss the Italian, German and Polish indexes, starting with the analysis of extreme air pollution events and moving toward median situations, both in spatial and in toxicity terms. Note that, for making the results interpretation easier, in each plot the index time series is integrated or replaced by a *Loess* curve curve computed using a smoothing parameter equal to 0.3.

4.1 Analysis of extreme pollution

For evaluating air pollution extreme values, we plot the *Maxmax*index of equation (1) using a different point style according to the pollutant that, at the last aggregation level, corresponds to the maximum.

	Piedmont-Lombardy	Berlin-Brandenburg	Masovian province
PS.MM	1.91	1.01	1.33
SP.mM	0.47	0.67	0.66
PS.Mm	1.04	0.70	0.78
V	0.72	0.31	0.48

Table 8: Annual average of considered indexes.

4.1.1 Piedmont-Lombardy

With reference to the Italian case, Figure 4 shows that $I\,(PS.MM)$ is above unity for almost all the year and, from Table 8, we see that the average level, being 1.91, is the highest. Moreover, it can be seen that PM_{10} and O_3 are the most critical pollutants.

During summer, Ozone stands out and its concentration exceeds for three times the doubled threshold of 120 $\mu g/m^3$. In the rest of the year PM_{10} results to be the most dangerous pollutant with toxicity levels that increase in winter. Note that the smoothed values stay permanently above twice the standard limit for seven months a year, moreover for 40 days, maximum daily PM_{10} concentrations are more than 3 times the limit value. This severe situation is known in Italy and the local governments, with the declared objective of reducing emissions, are experimenting some programs, which range from temporary measures, such as traffic reductions and periodical blocks, to permanent ones, such as limitations for old cars and incentives for low emission cars.

Figure 5 gives the distribution of the number of times that each station results to be the worst. From the underlying analysis it results that *Itis Grassi* station, Turin, is the worst station and attains the maximum 48 days a year. The second worst station is Trezzo d'Adda, East of Milan Province, with 37 days.

4.1.2 Berlin-Brandenburg

The German case is plotted in Figure 6 and has an average of 1.00 which is quite lower then the Italian case and the minimum among the three regions

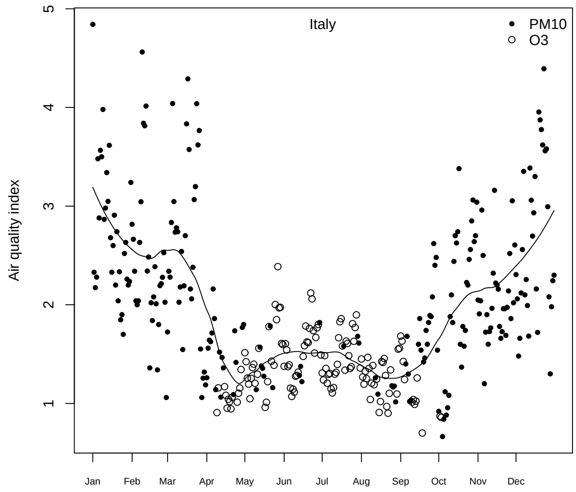


Figure 4: Italian air quality index $I_d\,(PS.MM)$ according to the decisive pollutant.

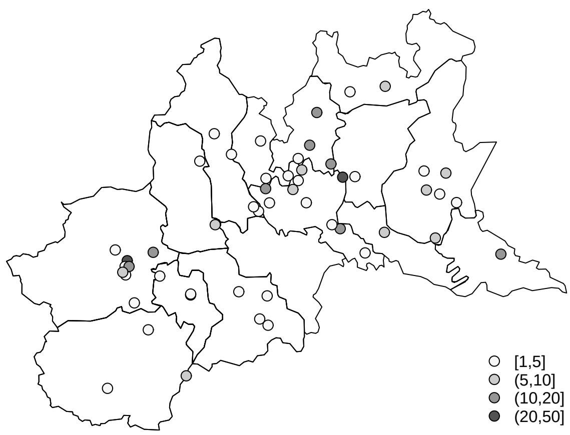


Figure 5: Italian worst station distribution.

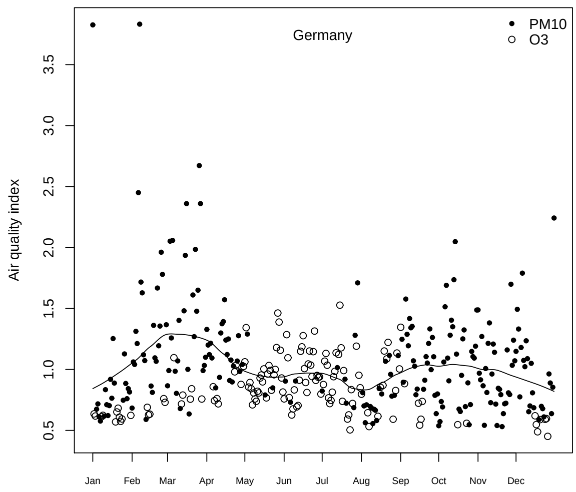


Figure 6: German air quality index $I_d\,(PS.MM)$ according to the decisive pollutant.

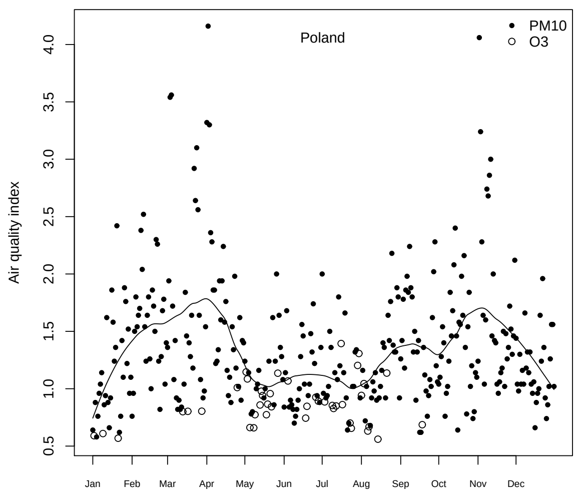


Figure 7: Polish air quality index $I_d\,(PS.MM)$ according to the decisive pollutant.

considered. The seasonal pattern of the worst pollutant is somewhat similar to the Italian one, as PM_{10} and O_3 are the worst pollutants in winter and summer respectively. The smoothed index is mainly between 0.5 and 1.0, except February and March when it is larger. The station analysis shows that the worst station is *Cottbus*, which is located near a traffic centre attaining the maximum for 90 days out of 365, the second worst station is located in Berlin. The two daily extreme observations also come from this station: the first one, as in Italy, is on January 1^{st} and can be explained by the New Year's fireworks.

4.1.3 Masovian Province

In Poland the average of the Maxmax index, being 1.33, is intermediate among the other two regions above. Its seasonal behavior, which is reported in Figure 7, is similar to the German one as it lacks the summer peak which is typical of the Italian pattern. Nevertheless, the main incisive pollutant is almost always PM_{10} and the seasonality is more pronounced then Berlin-Brandenburg with marked peaks in autumn and spring. The worst pollution results are observed in the two stations located in Warsaw, where the maximum values of the index are respectively measured for 110 and 63 days.

Another consideration regards the role of ozone in Polish data, which, differently form the other two regions, is scarcely the decisive pollutant, even in summer. As only eight stations out of 21 are equipped with ozone sensors this effect may be due to unbalanced network design and will be considered further in section 4.2.1.

4.1.4 Comparisons

The comparison of extreme pollution for year 2005 in the three regions shows that they are quite different not only for the yearly average, which is quite higher in North of Italy and lower in Berlin-Brandenburg, but also for the seasonal pattern.

In particular, the Italian index is characterized by a strong seasonality with a larger peak in winter, when PM_{10} is the main cause of high pollution, and a secondary peak in summer, when O_3 is the main hazard for humans and environment. On the other side, in the German and Polish regions PM_{10} has two different peaks, one in early spring and the other in autumn. Moreover the summer peak is almost absent for both regions.

The difference in the yearly average is consistent with the general higher anthropic pressure in the Northern Italian regions which interacts with the climatic component. As a matter of fact, the difference in the summer peaks is enlarged by the difference in solar radiation which amplifies the Italian ozone summer peak.

Moreover the one-winter-peak pattern in Italy is related to the long periods of weather stability which are common in December and January and is different from the North European pattern of Berlin where autumn and spring are more stable and dryer seasons favoring a moderate pollution accumulation.

In terms of pollution severity, 96% of year 2005, the Piedmont-Lombardy index $I\,(PS.MM)$ exceeded the unit standard limit value, whilst the same percentage for the German and Polish regions was 68% and 42%. It follows that in Northern Italy extreme toxic events are more likely to occur and that in Berlin-Brandenburg we observe the less severe pollution situation. The Masovian region pollution is intermediate but we have an additional uncertainty related to a sparser monitoring network.

4.2 Analysis of median pollution

In this section, we consider the use of the two aggregating strategies for the median indexes of equations (2) and (4) . The first one can be recommended for balanced networks and its capability of understanding spatial variability is illustrated. Vice versa, the second one results to be more stable or robust with respect to unbalanced multisensor network designs.

4.2.1 Spatial median of the worst pollutant

For analyzing the spatial median of the worst pollutant and its temporal dynamics, we use index $I\,(SP.mM)$ of equation (2) together with the dispersion measure V given by equation (3).

For the Italian case, which is plotted in Figure 8, it results that the index is always lower than one, with an average given by Table 8, which is the lowest of the three regions and contradicts the previous section conclusions. Such a bias follows from the unbalanced design of the Italian network which has only 79 PM_{10} sensors out of 199 stations. This network design bias is also suggested by the high values of spatial, or network, variability index V which is the highest for the Italian data.

As shown by Figures 9 and 10 and Table 8, the German and Polish data give lower spatial or network dispersion V, especially for Berlin-Brandenburg. As a matter of fact, index $I\,(SP.mM)$ has a behavior which is closer to the Maxmax index of previous section. Here the average of $I\,(SP.mM)$ is slightly lower for Polish data than Berlin-Brandenburg. Once again, we disregard the result as the dispersion V in Masovian region is quite higher than the German data, suggesting that the Masovian network is more unbalanced than the German one for the $I\,(SP.mM)$ index.

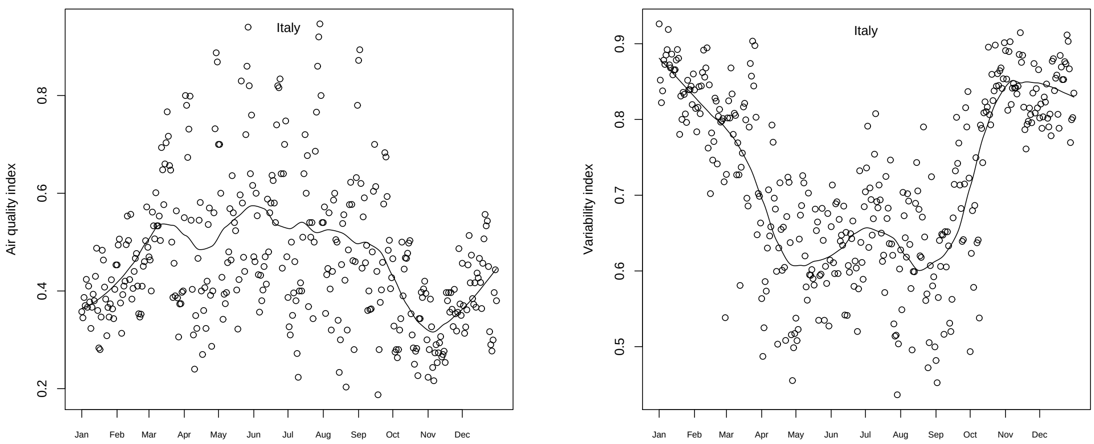


Figure 8: Italian air quality index $I_d\,(SP.mM)$ (left side) and dispersion measure V_d (right side).

To reinforce the conclusion that the high values of Italian V are related to network design rather than genuine spatial variability, we performed the same analysis for some representative provinces out of the 19 single provinces of Piedmont-Lombardy. We do not report the detailed figures here for the sake of brevity, nevertheless the results are essentially the same as the aggregate level. In particular, we get high values for V even at the province level confirming the idea of network heterogeneity by design.

4.2.2 Worst median pollutant

The second approach to median pollution is based on the index $I\,(PS.Mm)$ of equation (4). As it takes the median among the stations for each pollutant and then the maximum among the pollutants, it attenuates the dependence of the index on the network multisensor design.

Looking at Figure 11 and Table 8, it can be seen that, the Italian index $I\,(PS.Mm)$, has an average of 1.04 and differs markedly from index $I\,(SP.mM)$ both in average and seasonality. On the other hand it has a seasonal pattern similar to the Maxmax index $I\,(SP.MM)$ of Figure 4.

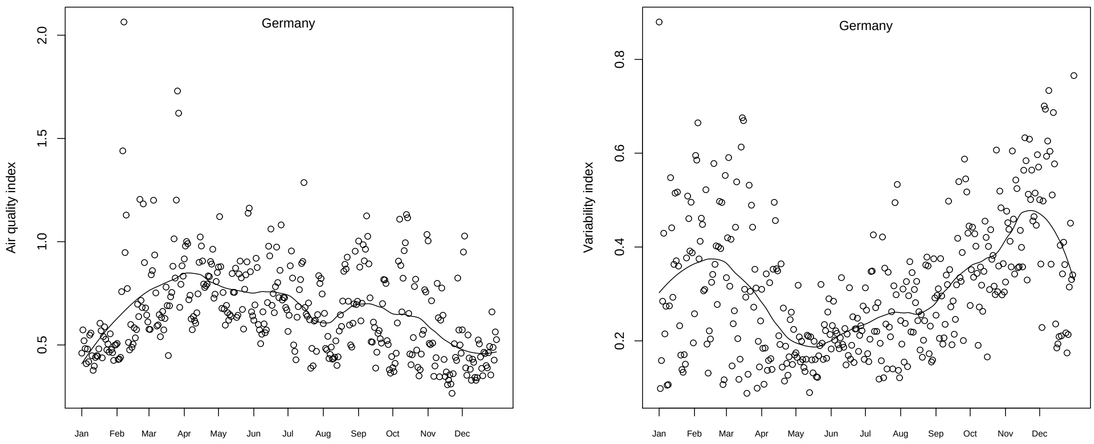


Figure 9: German air quality index $I_d(SP.mM)$ (left side) and dispersion measure V_d (right side).

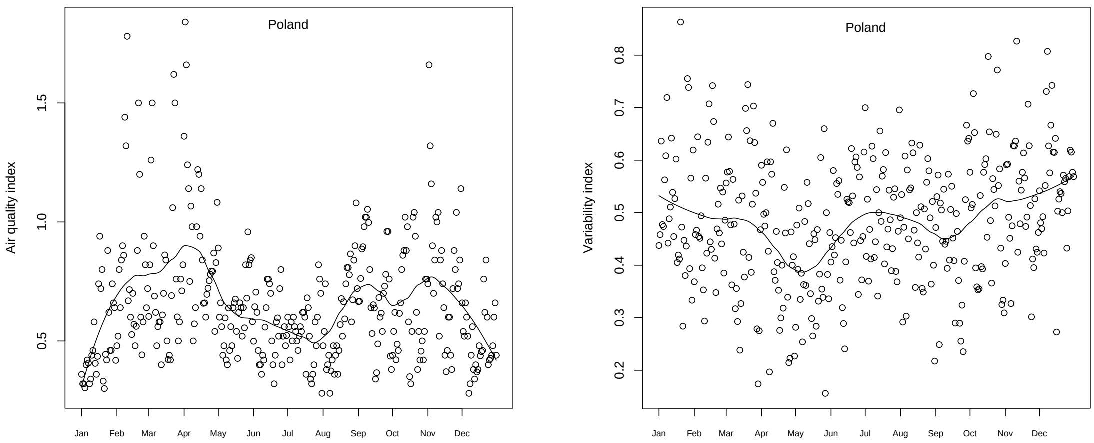


Figure 10: Polish air quality index $I_d(SP.mM)$ (left side) and dispersion measure V_d (right side).

For the German case, Figure 11 shows that there are no remarkable differences between $I\,(PS.Mm)$ and $I\,(SP.mM)$ neither in the scale nor in the shape. For the Polish case of Figure 12, instead, it is worth to note that this index is slightly greater than $I\,(SP.mM)$, especially in summer.

Hence, the $Maxmax$ analysis of section 4.1 is confirmed by the median analysis of index $I\,(PS.Mm)$. Moreover the latter has to be preferred for describing the median pollution with respect to $I\,(SP.mM)$, as it does not loose information about the average level and the seasonal pattern in case of unbalanced networks.

4.2.3 Quantile comparisons

The right side of Figure 12 refers to the empirical distribution function of the worst median pollution index of previous section and can be used for prompt index comparisons. For example, the severe air pollution conditions of Piedmont-Lombardy results in being for 47.9% of the year above the limit values; whereas for the other two North European regions, Berlin-Brandenburg and Masovian Province, this happens for 9.3% and 15.8% of the year, respectively. This and the non overlapping behavior of the three distribution functions confirms the fact that the best air quality situation is referred to the Berlin-Brandenburg area.

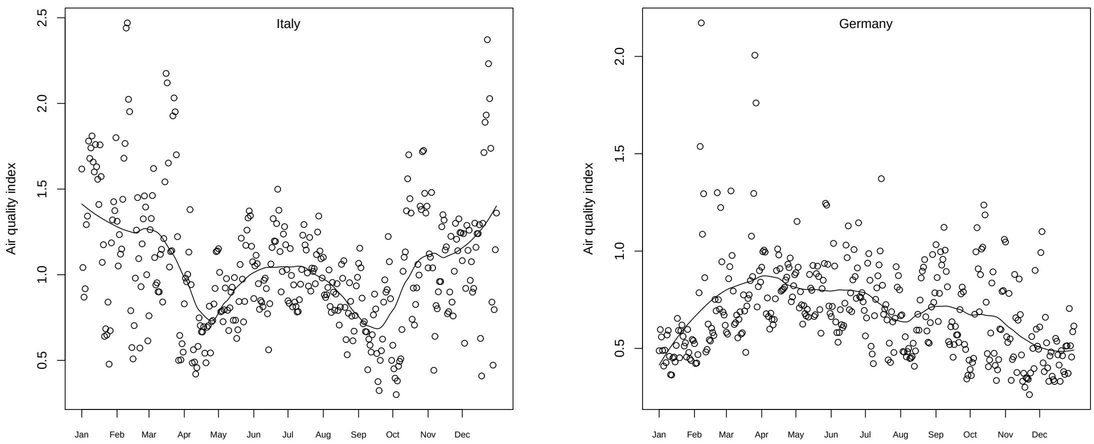
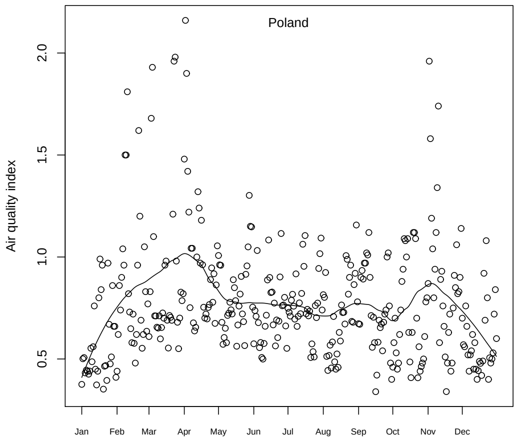


Figure 11: Left side: Italian $I_d\,(PS.Mm)$ index. Right side: German $I_d\,(PS.Mm)$ index



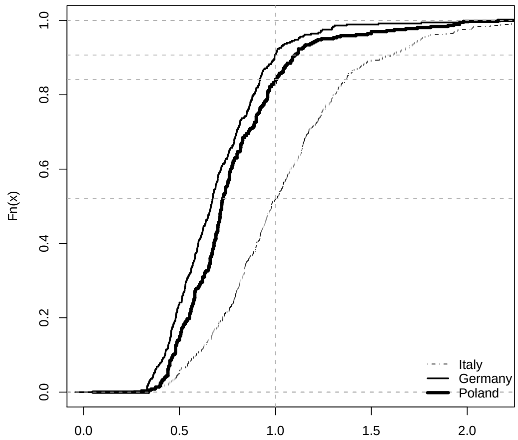


Figure 12: Left side: Polish $I_d\,(PS.Mm)$ index. Right side: empirical distribution functions for $I_d\,(PS.Mm)$.

5 Conclusions

In this work, in the perspective of defining an European common index methodology, which makes air quality comparable in time and across different countries, we analyzed the behavior of BC indexes for comparing air pollution in three different European regions.

To see this, we showed how to use the BC indexes for synthetic description and communication of daily global pollution and for regional comparisons.

Moreover, we highlighted the interplay between the monitoring network structure and the index behavior. Thanks to this, we showed that the BC index may be useful to understand the network structure and vice versa, knowing the network structure gives guidance to the index to be used.

In particular it turns out that the BC index based on the spatial median of the maximum among pollutants of each station, denoted by $I\,(SP.mM)$, may be used for describing and comparing the mean pollution if the network is balanced. This index may be coupled with a spatial dispersion index for assessing the between stations variability and the balanced network hypothesis.

Moreover, we showed that two indexes, namely the Maxmax index and the worst median pollutant index, which is denoted by $I\,(PS.Mm)$, are more robust with respect to network design and can be used to describe and compare different regions. In particular, they highlight various properties of daily pollution, such as the particular seasonality behavior of Northern Italy, which is characterized by different pollutants in different seasons and winter and summer peaks.

6 Acknowledgement

The authors would like to thank the Umweltbundesamt (Fachgebiet II 5.2, Immissionssituation), the Landesumweltamt Brandenburg (Abt. Immissionsschutz), the Berliner Luftgüte Messnetz (BLUME) and the Chief Inspectorate of Environmental Protection (Department of Monitoring, Assessment and Outlooks, Warsaw) for the given support and help.

The work of M. Cameletti and A. Fassò was partially supported by PRIN project n.2006131039 and Regione Piemonte CIPE project 2004.

References

Bellini, P., Baccini, M., Biggeri, A., and Terracini, B. (2007). The meta-analysis of the Italian studies on short-term effects of air pollution (MISA): old and new issues on the interpretation of the statistical evidences. *Environmetrics*, **18**, 219–229.

Bruno, F. and Cocchi, D. (2002). A unified strategy for building simple air quality indexes. *Environmetrics*, **13**, 243–261.

Bruno, F. and Cocchi, D. (2007). Recovering information from synthetic air quality indexes. *Environmetrics*, **18**, 345–359.

Chiu, G., Guttorp, P., and Westveld, A. (2007). A latent health factor index via generalized linear mixed models, with application to ecological health assessment. Submitted to Applied Statistics.

Englert, N. (2004). Fine particles and human health — a review of epidemiological studies. *Toxicology Letters*, **149**, 235–242.

European Environment Agency (2007). *How can I protect my health?* http://www.eea.europa.eu/maps/ozone/whatcanIdo.

Fassò, A. and Negri, I. (2002). Nonlinear statistical modelling of high frequency ground ozone data. *Environmetrics*, **13**, 225–241.

Fassò, A., Cameletti, M., and Nicolis, O. (2007). Air quality monitoring using heterogeneous networks. *Environmetrics*, **18**, 245–264.

Lagona, F. (2005). Air quality indices via non homogeneous hidden markov models. In *Proceedings of the Italian Statistical Society Conference on Statistics and Environment, Contributed Papers, CLEUP - Padova, 91-94*.

Peel, M. C., Finlayson, B. L., and McMahon, T. A. (2007). Updated world map of the Köppen-Geiger climate classification. *Hydrology and Earth System Sciences*, **11**, 1633–1644.

Pope, C. (2000). Review: Epidemiological basis for particulate air pollution health standards. *Aerosol Science and Technology*, **32**, 4–14.

Saisana, M., Saltelli, A., and Tarantola, S. (2005). Uncertainty and sensitivity analysis techniques as tools for the quality assessment of composite indicators. *Journal of the Royal Statistical Society A*, **168**, 307–323.

U.S. EPA (2003). Guideline for the reporting of daily air quality - air quality index. http://airnow.gov/index.cfm?action=static.publications.