The Quality of Urban Air in Barcelona: A New Approach Applying Compositional Data Analysis Methods

J. Gibergans-Báguena [a*], C. Hervada-Sala [a], E. Jarauta-Bragulat [b]

[a] U. Politècnica de Catalunya - BarcelonaTech (UPC), ESEIAAT, Colom 1, 08222. Terrassa, Spain

[b] U. Politècnica de Catalunya - BarcelonaTech (UPC), ETSECCPB, Jordi Girona 1-3, 08034. Barcelona, Spain

Abstract

The main goal of this paper is to go some steps further to improve the understanding and manageability of air quality. Quality of atmospheric air in large cities is a matter of great importance because of its impact on the environment and on the health of the population. Recently, measures restricting access of private vehicles to the centre of large cities and other measures to prevent atmospheric air pollution are currently topical. The knowledge of air quality acquires special relevance to be able to evaluate the impact of those great social and economic measures. There are many indices to express air quality. In fact, quite every country has its own, depending on the main pollutants. In general, all indices ignore the compositional nature of the concentrations of air pollutants and do not apply methods of Compositional Data Analysis and have some other weak points such as leak of standardized scale. Therefore, the methodology used is founded on Compositional Data Analysis. The air quality index has an adequate correlation between input (concentrations) and output (air quality index), it distinguishes between air pollution and air quality and it has a 0-100 reference scale which makes easier interpretation and management of air quality expression. To illustrate the proposed method, an application is made to a series of air pollution data (Barcelona, 2001-2015). The results show the effectiveness of the 2008 European directive on ambient air quality.

1- Introduction

Air pollution in cities, mainly in large cities or densely populated areas, is a burning issue that concerns citizens because of its impact on daily life and its consequences on people's health [1-6]. More than a half of the planet's inhabitants live in urban areas. This is the reason why in large cities the quality of the environment in general and the air in particular is a problem that deserves special attention. Implications of atmospheric air pollution go far beyond the health of people and affect the environment in general, the economy and the future of life on our planet [7]. Thus, the increasing demand on air quality makes it a point in urban management: cities must take measures to guarantee that air quality is at adequate levels to avoid affections on the health of population. It is very important, therefore, to have an adequate methodology to quantify the expression of atmospheric air quality to help decision makers to control it correctly.

Atmospheric air pollution is usually expressed by a numerical value called "Air Quality Index" (AQI). This index is obtained from concentrations of some air pollutants, usually: O3, CO, NO2, SO2 and suspended particulate matter of certain size or diameter: lower than 2.5 microns (PM2.5) and lower than 10 microns (PM10). Presence of pollutants and particulate matter is expressed by their concentration in units of mass relative to a total volume unit of air usually $\mu g/m^3$. There are several methodologies to express the quality of atmospheric air from the concentrations of air

pollutants. The best known is the proposal by the USA Environmental Protection Agency (EPA), which is based on a piecewise linear function that transforms the concentrations into AQI values in a certain scale. However, all the methodologies applied until now, handle concentrations of air pollutants ignoring their compositional nature and therefore committing some errors, such as calculating arithmetic averages [8-10].

Although there are other authors who use methodology that has to do with indexes [11, 12], we believe that our proposition based on compositional data is more consistent with the nature of the kind of data under study. The most recent proposal for an Air Quality Index that takes into account the compositional nature of the data is, as far as we know, the one developed in [10]. The proposal is based on the concept of logcontrast and an air quality index is defined (AQI*) as a function of the geometric mean of concentrations of six air pollutants (O3, NO2, CO, SO2, PM10, PM2.5). The index is scaled from zero to 100 using a proportionality factor, according to concentration's values.

In the present work, an improvement of that model is proposed, taking into account the compositional nature of the air pollutants concentrations, that is, applying Compositional Data Analysis methods. One of the purposes is giving an index that makes it clear that air pollution and air quality mean opposite things, which seems not so clear in most of the existing AQI. At the same time, this index establishes a different slope variation (derivative) in the low pollution zone and the high pollution zone, which allows for a better discrimination in low polluted areas. In addition to a global index of air quality, the proposal is giving an individual index for each of the pollutants, which makes it possible the detection of possible dangerous individual pollutant levels, that otherwise could keep unnoticed in a global index. At last, to help decision makers using it in a reliable, simple and adequate way, the new index has a natural scale to express the values of air quality. Figure 1 shows the improvements offered by our proposed air quality index.
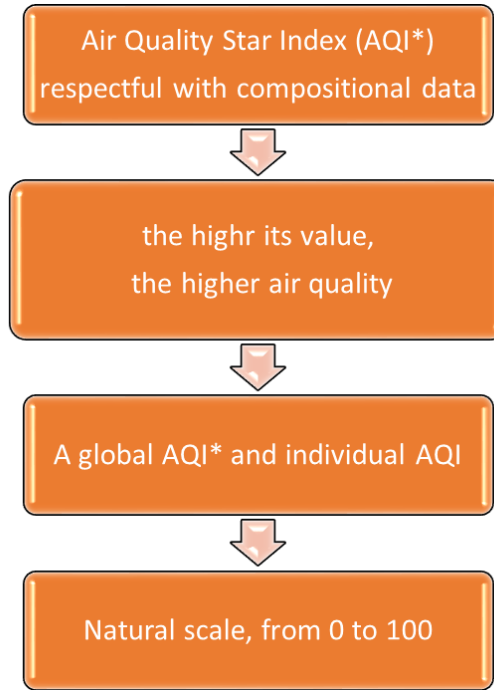

Figure 1. Improvements Air Quality Index.

2- Definition of Air Quality Star Index

Air pollution data are given usually as a real coefficient (N, D+1)-matrix M, where D stands for the number of pollutants and +1 for the period time. Therefore, its k-row can be expressed as:

$$M(k) = \begin{pmatrix} t_k & c_1(t_k) & c_2(t_k) & \cdots & c_D(t_k) \end{pmatrix}, \ k = 1, 2, ..., N \tag{1}$$

Being t_k the k-time period (day, week, month, …), k = 1,2,…,N, N the number of time periods and the concentration of ith air pollutant at time tk (units are usually μg/m3). The most significant air pollutants for their impact on people health in urban surroundings are O3, NO2, PM10 and PM2.5 [13-15]. Therefore, in this work we will consider four air pollutants (D = 4) and other components of air grouped in the so-called "residual component". Then, the k-th row we take is

$$M(k) = \begin{pmatrix} t_k & c_{O3}(t_k) & c_{NO2}(t_k) & c_{PM10}(t_k) & c_{PM2.5}(t_k) \end{pmatrix}. \tag{2}$$

A logcontrast (LC) is defined as a linear combination of logarithms of parts with coefficients adding up to zero [16]. If a logcontrast is computed considering air pollutants and the residual component, for any time t_k, k=1,2,…,N, put, then taking into account properties of logarithmic function, the equation we have can be written as:

$$LC = \log\left(c_{O3}{}^{\alpha_{O3}} c_{NO2}{}^{\alpha_{NO2}} c_{PM10}{}^{\alpha_{PM10}} c_{PM2.5}{}^{\alpha_{PM2.5}}\right) - \log\left(c_{res}\right) \tag{3}$$

$$\alpha_{O3} + \alpha_{NO2} + \alpha_{PM10} + \alpha_{PM2.5} = 1.$$

The filling-up value of air residual component is almost never reported, its computation is very difficult and the second term in the right-hand of Equation 3 is almost constant ([10]). Therefore, air pollution can be expressed using the considered air pollutants with Equation 3 or its approximation

$$f\left(c_{O3}, c_{NO2}, c_{PM10}, c_{PM2.5}\right) = c_{O3}{}^{\alpha_{O3}} c_{NO2}{}^{\alpha_{NO2}} c_{PM10}{}^{\alpha_{PM10}} c_{PM2.5}{}^{\alpha_{PM2.5}} \tag{4}$$

$$\alpha_{O3} + \alpha_{NO2} + \alpha_{PM10} + \alpha_{PM2.5} = 1.$$

From Equation 4, a function for an Air Pollution Index (API) calculation can be stated as:

$$API = K_{global}\left(c_{O3}{}^{\alpha_{O3}} c_{NO2}{}^{\alpha_{NO2}} c_{PM10}{}^{\alpha_{PM10}} c_{PM2.5}{}^{\alpha_{PM2.5}}\right)^{\beta}$$

$$K_{global} = \frac{100}{f\left(O3_{max}, NO2_{max}, PM10_{max}, PM2.5_{max}\right)} \tag{5}$$

$$\alpha_{O3} + \alpha_{NO2} + \alpha_{PM10} + \alpha_{PM2.5} = 1.$$

The exponents can be used to highlight the impact of an air pollutant in particular, according to its impact on the population health or other criteria. In this paper, the values adopted are ([13]):

$$\alpha_{O3} = 0.2, \ \alpha_{NO2} = 0.3, \ \alpha_{PM10} = \alpha_{PM2.5} = 0.25. \tag{6}$$

The β exponent allows applying for a nonlinear model and, thus, having a different shape of the pollution curve in the low pollution zone with respect to the high pollution zone. The value of the exponent used in this work is β = 1.25, as it has shown to be the best discriminant (see Figure 2).
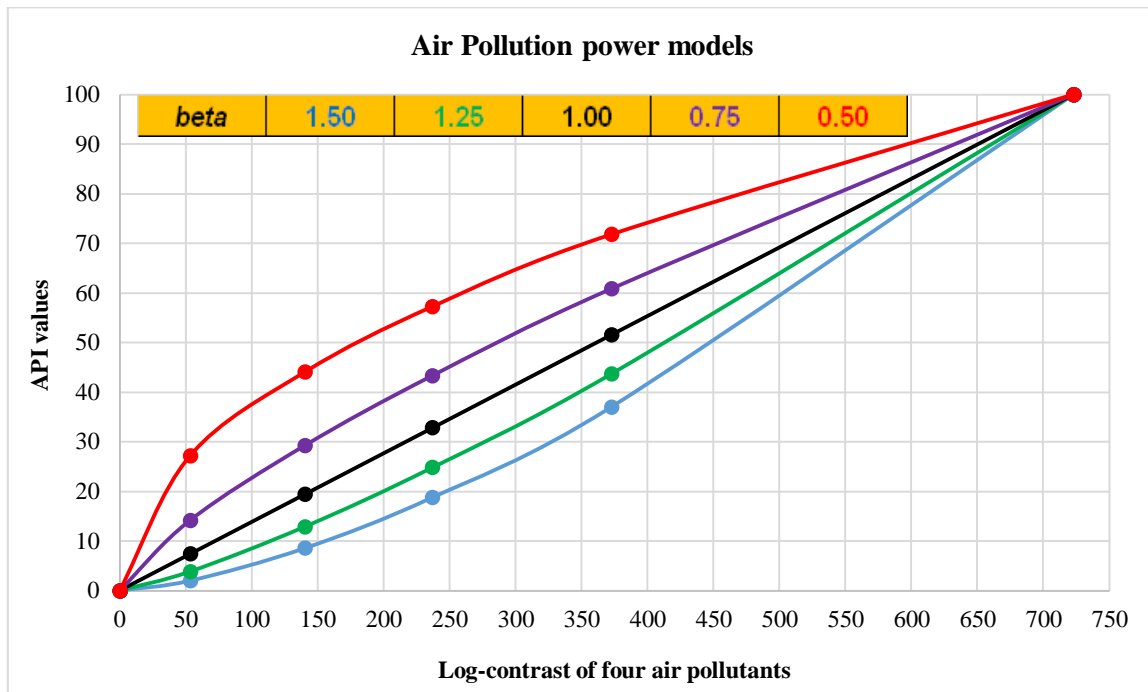

Figure 2. Different Air Pollution power models, computed from a log-contrast of four air pollutants. (API-1, power = 1.50; API-2, power = 1.25; API-3, power = 1.00 (linear model); API-4, power = 0.75 and API-5, power = 0.50).

The multiplicative factor K_{global} is introduced for ranging air pollution index values in a scale from zero to 100. The value of the multiplicative factor K_{global} depends on the value of β exponent, as well as on the maximum pollutant' concentrations adopted as maximum admissible pollution [14, 17]. An example is shown in Table 1 and the corresponding value of K_{global} is 100/223.8.

Thus, if considered air pollutants are O3, NO2, PM10 and PM2.5, the air pollution index expression is then

$$API = 0.0267\left(c_{O3}^{0.2}\; c_{NO2}^{0.3}\; c_{PM10}^{0.25} c_{PM2.5}^{0.25}\right)^{1.25} \tag{7}$$

Observe that API is increasing when air pollutant concentrations increase, that is, the trend of the values of this index is consistent with its name.

Once API has been defined, the definition of the Air Quality Index becomes simple:

$$AQI^* = 100 - API \tag{8}$$

Observe that AQI is decreasing when API is increasing, that is, it corresponds to the idea that to lower value of the AQI*, the lower quality of the atmospheric air, and vice versa, according to the intuitive significance of the word "quality".

The air pollution index calculated with those pollutants is an overall index; an average value that might not reflect possible episodes of high values of a given pollutant. Therefore, it is interesting to be able to calculate individual index. Thus, for a given air pollutant its own pollution index is defined through:

$$
\begin{aligned}
API_{O3} &= 0.0251\left(c_{O3}^{\;\alpha_{O3}}\right)^{1.25}; \; AQI*_{O3} = 100 - API_{O3} \\
API_{NO2} &= 0.0075\left(c_{NO2}^{\;\alpha_{NO2}}\right)^{1.25}; \; AQI*_{NO2} = 100 - API_{NO2} \\
API_{PM10} &= 0.0520\left(c_{PM10}^{\;\alpha_{PM10}}\right)^{1.25}; \; AQI*_{PM10} = 100 - API_{PM10} \\
API_{PM2.5} &= 0.1004\left(c_{PM2.5}^{\;\alpha_{PM2.5}}\right)^{1.25}; \; AQI*_{PM2.5} = 100 - API_{PM2.5}
\end{aligned}
\tag{9}
$$

Then using all information stated above, we use a colour code when giving AQI* values. Table 1 shows colour codes for individual AQI* values as well as for global, and their concentration breakpoints.

Table 1. Concentration breakpoints for each considered air pollutant; in bold, maximum values assigned. Column *f* indicates the result according to Equation 4 and exponents are at the last row. In last column, global AQI* values and corresponding color codes.

| O3 | | NO2 | | PM10 | | PM2.5 | | *f* | Global AQI* |
µg/m³	AQI*	µg/m3	AQI*	µg/m³	AQI*	µg/m3	AQI*		
0.0	100	0.0	100	0.0	100	0.0	100	0.0	100
24.0	89	40.0	99	20.0	92.4	15.0	96.9	22.4	96
96.0	84.5	160.0	92.5	80.0	71.8	60.0	89.8	89.5	87
156.0	79.8	260.0	77.7	130.0	47.3	97.5	81.3	145.4	74
204.0	74.8	340.0	55.4	170.0	20.2	127.5	47.1	190.2	52
240.0		**400.0**		**200.0**		**150.0**		**223.8**	**0.0**

α_i	0.25		0.20		0.30		0.25	

3- Application to Barcelona Urban Area 2001-2015

Barcelona is a city and metropolis on the Mediterranean coast of the Iberian Peninsula. It is the capital of Catalonia, and the second city in population and economic weight of the Iberian Peninsula, after Madrid.

3-1- Data Set

The data set consists of a monthly data matrix from 2001 to 2015, which is 180 months. Data have been imported directly from the European Environmental Agency (EEA's Central Data Repository) [18] where the country reports are stored. The data in the web-based application (AIDE D) reflect the current/live status of the latest data, uploaded by countries and successfully tested. The average monthly air pollutants O3, NO2, PM10 and PM2.5 are available. The value of the monthly average was computed as the geometric mean of the monthly mean values of 18 air data measuring stations located in the city of Barcelona, as indicated in Figure 3.
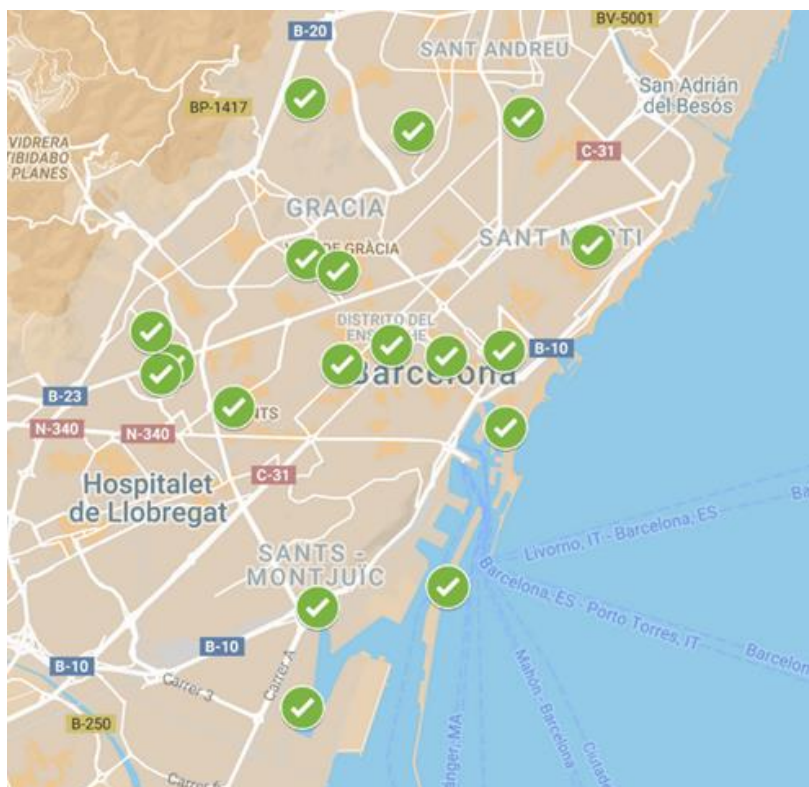
Figure 3. Map of Barcelona, showing the location of air pollution measuring points (green).

In Table 2, there are the minimum, maximum and geometric mean values of O3, NO2, PM10 and PM2.5 concentration in µg/m^3 in Barcelona. Figure 4 shows pollutant's concentration boxplot.

Table 2. Concentrations in µg/m^3.

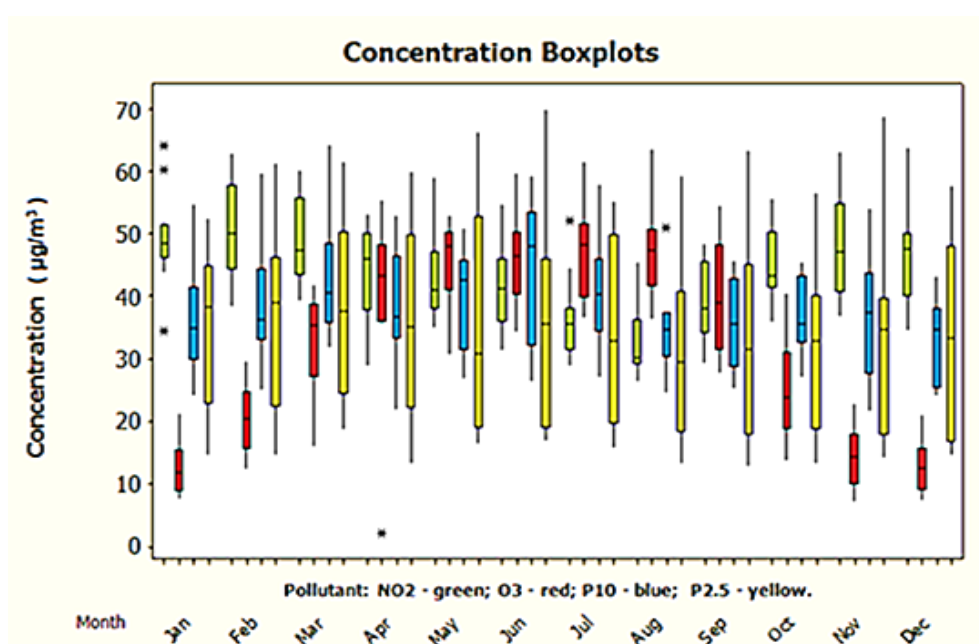|	Ozone	NO2	PM10	PM2.5
Geometric mean	28.39	42.26	33.95	25.81
Minimum value	2.18	25.34	17.46	11.17
Màximum value	63.58	64.10	64.36	70.00

Figure 4. Pollutants' concentration boxplot in Barcelona.

3-2- How to Apply AQI*

As there is stated that from the point of view of compositional data analysis, the use of log-contrasts is much more suitable and coherent as using concentrations or even log-concentrations, Figure 5 shows Log contrast and API evolution:
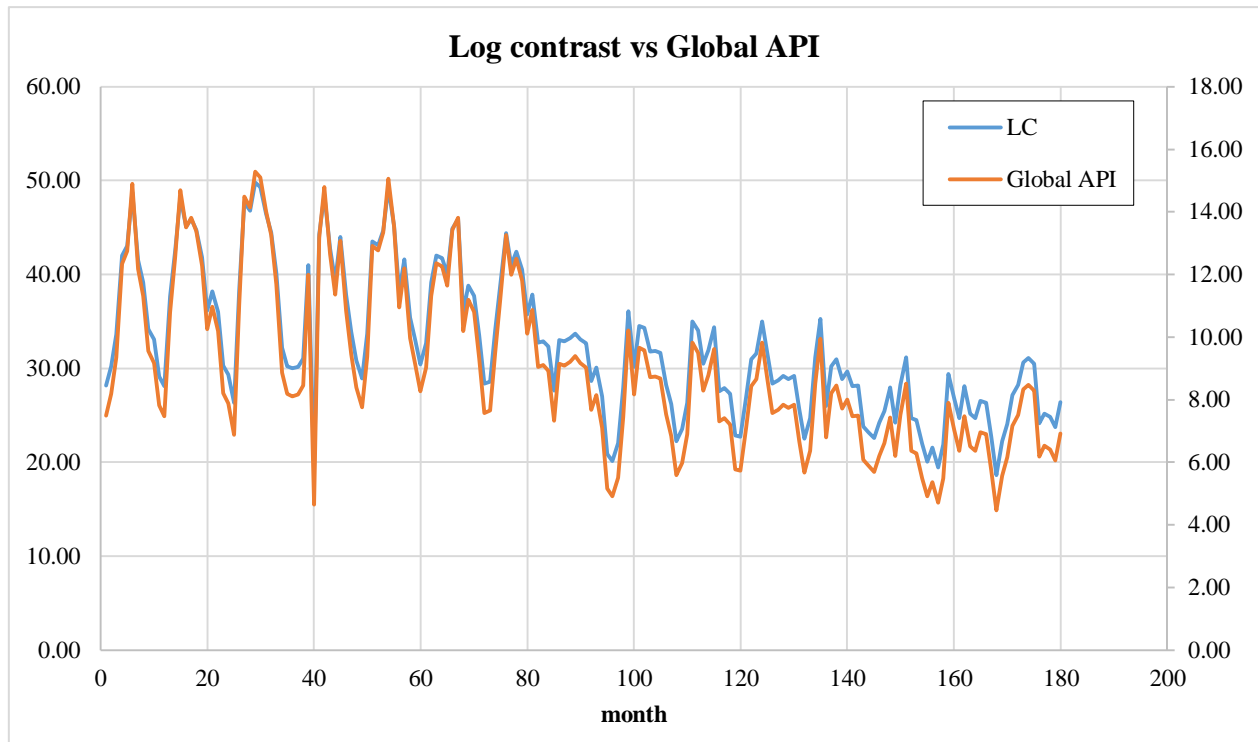

Figure 5. Log contrast (LC) and Air Pollution Index (API) evolution in Barcelona from 2001 to 2015.

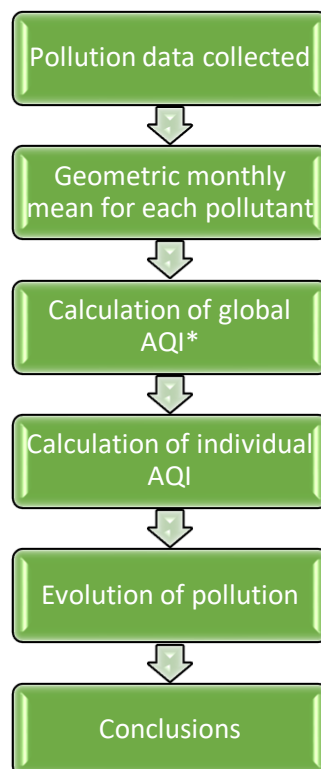Figure 6 is shows how to calculate AQI* from the whole data set.

Figure 6. Flowchart of methodology to calculate AQI*.

In Figure 7 global AQI* function is plotted. For this period, average of global AQI* values is 90.95, a reasonably good value for air quality. One can observe an increasing air quality value by January 2009 (red mark).
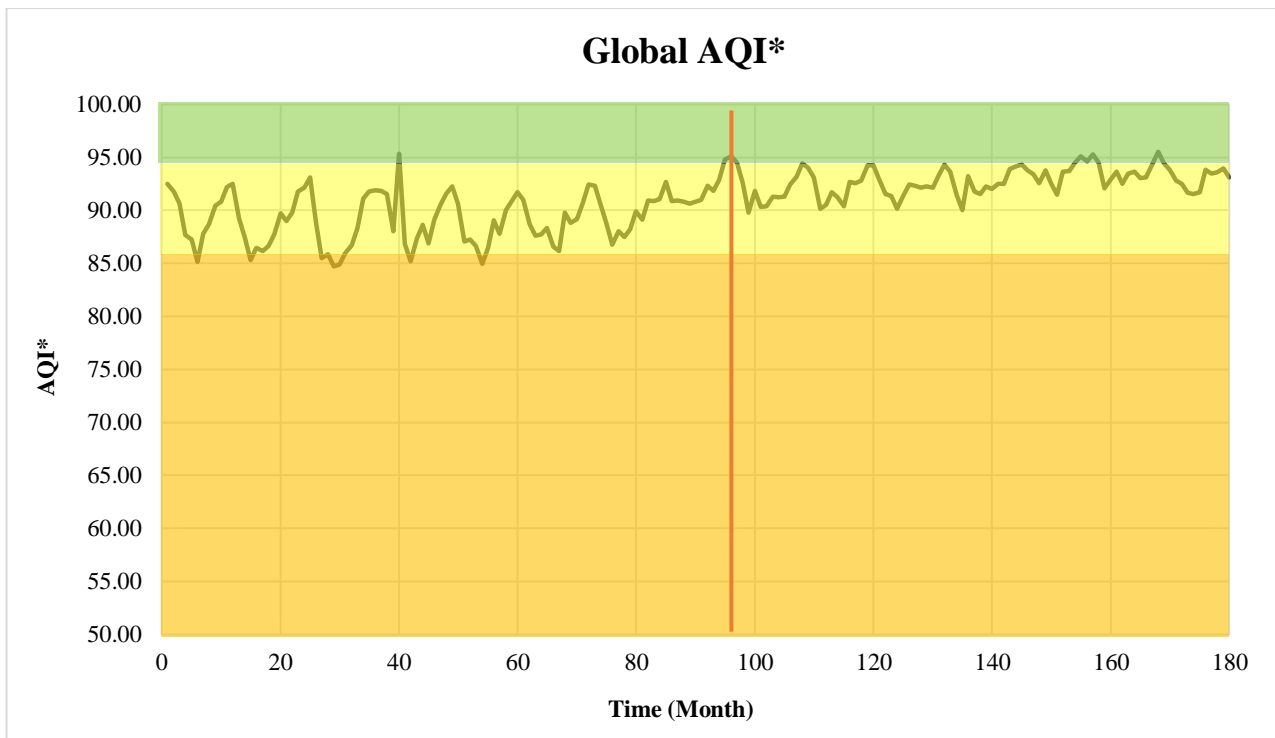

Figure 7. Global AQI* function. Data series: Barcelona 2001-2015. Red line stands for January'09.

Figure 8 shows AQI for each pollutant. Average individual AQI* values are 91.36 for O3, 93.80 for NO2, 88.40 for PM10 and 86.22 for PM2.5.There is also an increase of air quality by January 2009 (red mark).
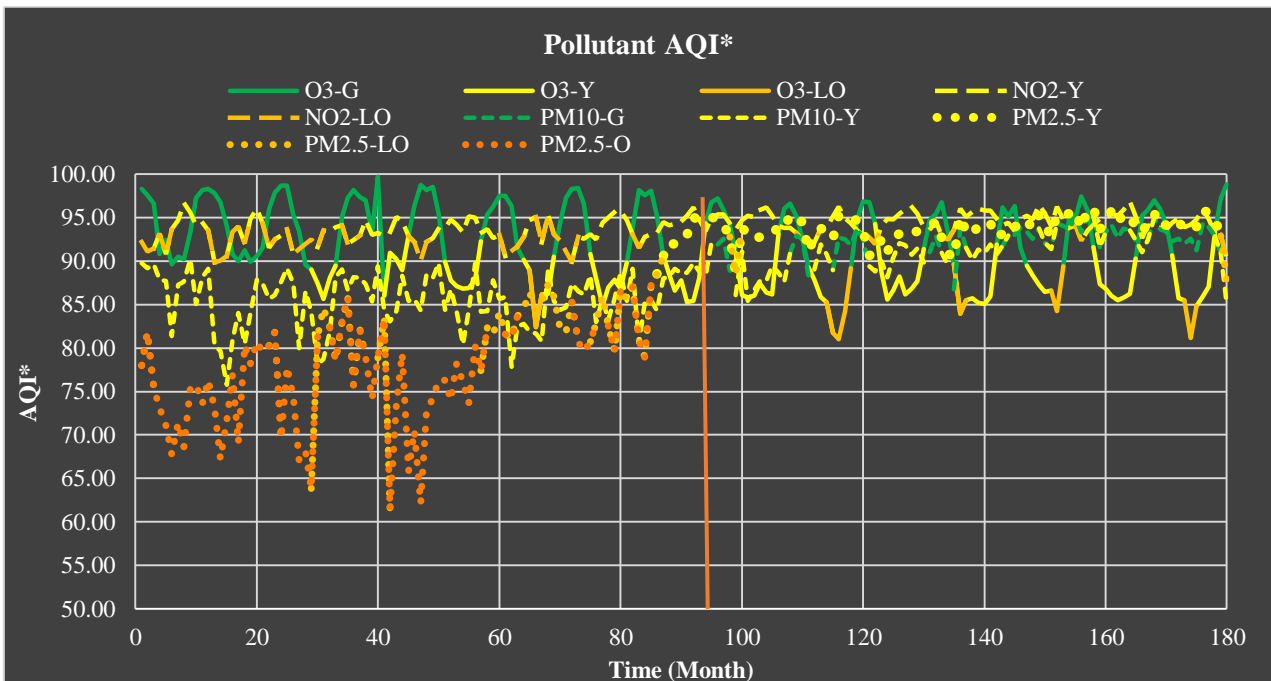

Figure 8. Individual AQI* functions for air pollutants. Data series: Barcelona 2001-2015. Red line stands for January'09.

Table 3 shows the percentage of months with a given air quality, for each pollutant as well as the global one. It is clearly observable that global air quality increases in Barcelona after 2009, mainly for the increase of AQI for small particles. The European directive on clean air, approved in May 2008, can induce this. However, it seems to have no effect on ozone concentration.

Table 3. Percentage of months with a given air quality in Barcelona.

2001-2008	O3	NO2	PM10	PM2.5	GLOBAL	2009-2015	O3	NO2	PM10	PM2.5	GLOBAL
	100,0	43,7	93,7				100,0	75,0	100,0	18,7	
		56,3	6,3	75,0	79,2			25,0		81,3	100,0
				25,0	20,8						

4- Conclusions

Four main conclusions can be stated from this work:

- The traditional methodology on which the expression of atmospheric air quality is based, presents certain inaccuracies and deficiencies. Compositional data analysis provides a more appropriate conceptual and methodological framework for establishing a quantitative air quality index based on air pollutant concentrations. The concept of log-contrast is key to developing the new methodology.

- The new air quality index proposed in this paper seems more appropriate to express air quality because it improves the correlation between input and output, establishes a reference scale and it is based on a methodology that takes into account the nature of the numerical data (compositional data).

- The air quality must be reported using an Air Quality Report (AQR). This AQR must have the overall air quality index AQI* and individual AQI for those air pollutants which quality is lower than the corresponding breakpoint.

- Attending to air quality in Barcelona, 2001-2015: (a) Global air quality is satisfactory and has improved in the last years. (b) The new proposed AQR methodology makes it possible to express more accurately air quality in Barcelona and it allows detecting episodes of possible risk for the population that could not be detected using the traditional air quality expression. (c) The impact on air quality is mainly due to the presence of particles rather than the presence of gases. (d) Pollution from particles has reduced mainly from 2009 onwards. Thanks to the DIRECTIVE 2008/50/EC OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL [19], on ambient air quality and cleaner air for Europe. (e) The individual index of ozone has worsened in that period.

5- Funding and Acknowledgments

6- Conflict of Interest

The author declares that there is no conflict of interests regarding the publication of this manuscript. In addition, the ethical issues, including plagiarism, informed consent, misconduct, data fabrication and/or falsification, double publication and/or submission, and redundancies have been completely observed by the authors.

7- References

[1] Zivin, Joshua Graff, and Matthew Neidell. "Air pollution's hidden impacts." Science 359, no. 6371 (2018): 39-40. doi:10.1126/science.aap7711.

[2] Grzędzicka, Emilia. "Is the Existing Urban Greenery Enough to Cope with Current Concentrations of PM2.5, PM10 and CO2?" Atmospheric Pollution Research 10, no. 1 (January 2019): 219–233. doi:10.1016/j.apr.2018.08.002.

[3] Lehtomäki, Heli, Antti Korhonen, Arja Asikainen, Niko Karvosenoja, Kaarle Kupiainen, Ville-Veikko Paunu, Mikko Savolahti, et al. "Health Impacts of Ambient Air Pollution in Finland." International Journal of Environmental Research and Public Health 15, no. 4 (April 12, 2018): 736. doi:10.3390/ijerph15040736.

[4] Kerimray, Aiymgul, Bulat Kenessov, and Ferhat Karaca. "Trends and Health Impacts of Major Urban Air Pollutants in Kazakhstan." Journal of the Air & Waste Management Association 69, no. 11 (October 9, 2019): 1331–1347. doi:10.1080/10962247.2019.1668873.

[5] Nehr, Sascha, Isabelle Franzen-Reuter, and Catharina Kucejko. "New Directions: Future Approaches to the Standardized Assessment of Airborne Pollutants Affecting Environmental Quality." Atmospheric Environment 166 (October 2017): 570–572. doi:10.1016/j.atmosenv.2017.06.051.

[6] Nowak, David J., Satoshi Hirabayashi, Marlene Doyle, Mark McGovern, and Jon Pasher. "Air Pollution Removal by Urban Forests in Canada and Its Effect on Air Quality and Human Health." Urban Forestry & Urban Greening 29 (January 2018): 40–48. doi:10.1016/j.ufug.2017.10.019.

[7] Hawking S. "Stephen Hawking says humans must colonize another planet in 100 years or face extinction". (2017) (Available online: http://www.cnbc.com/2017/05/05/stephen-hawking-human-extinction-colonize-planet.html.

[8] Hervada-Sala C, Gibergans-Báguena J, Jarauta-Bragulat E. "Speed limitation in metropolitan area of Barcelona: impact on air quality." Proceedings of International Conference on Air Quality, (2018). pp 49.

[9] Plaia, Antonella, and Mariantonietta Ruggieri. "Air Quality Indices: a Review." Reviews in Environmental Science and Bio/Technology 10, no. 2 (December 14, 2010): 165–179. doi:10.1007/s11157-010-9227-2.

[10] Jarauta-Bragulat, Eusebi, Carme Hervada-Sala, and Juan José Egozcue. "Air Quality Index Revisited from a Compositional Point of View." Mathematical Geosciences 48, no. 5 (May 23, 2015): 581–593. doi:10.1007/s11004-015-9599-5.

[11] Morawska, Lidia, Phong K. Thai, Xiaoting Liu, Akwasi Asumadu-Sakyi, Godwin Ayoko, Alena Bartonova, Andrea Bedini, et al. "Applications of Low-Cost Sensing Technologies for Air Quality Monitoring and Exposure Assessment: How Far Have They Gone?" Environment International 116 (July 2018): 286–299. doi:10.1016/j.envint.2018.04.018.

[12] Hao, Yan, and Chengshi Tian. "The Study and Application of a Novel Hybrid System for Air Quality Early-Warning." Applied Soft Computing 74 (January 2019): 729–746. doi:10.1016/j.asoc.2018.09.005.

[13] Pope, C. Arden, and Douglas W. Dockery. "Health Effects of Fine Particulate Air Pollution: Lines That Connect." Journal of the Air & Waste Management Association 56, no. 6 (June 2006): 709–742. doi:10.1080/10473289.2006.10464485.

[14] Zhang, Jing, Lan-yue Zhang, Ming Du, Wei Zhang, Xin Huang, Ya-qi Zhang, Yue-yi Yang, et al. "Indentifying the Major Air Pollutants Base on Factor and Cluster Analysis, a Case Study in 74 Chinese Cities." Atmospheric Environment 144 (November 2016): 37–46. doi:10.1016/j.atmosenv.2016.08.066.

[15] Van den Elshout, Sef, Karine Léger, and Hermann Heich. "CAQI Common Air Quality Index — Update with PM2.5 and Sensitivity Analysis." Science of The Total Environment 488–489 (August 2014): 461–468. doi:10.1016/j.scitotenv.2013.10.060.

[16] Aitchison J. "The Statistical Analysis of Compositional Data. Monographs on Statistics and Applied Probability, First Edition" (August 21, 1986), Chapman & Hall Ltd, London (UK).

[17] European Environment Agency. Air quality in Europe - 2018 report. EEA Report No 12/2018, (2018). pp 83.

[18] European Environment Agency. Air quality in Europe. Available online: https://www.eea.europa.eu/data-and-maps/data/.

[19] Official Journal of the European Union. 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. L 152/1. pp. 44.