The expression of air quality in urban areas: going further on a Compositional Data Analysis approach

J.Gibergans-Báguena[1], C.Hervada-Sala[1] and E.Jarauta-Bragulat[2]

[1] U. Politècnica de Catalunya - BarcelonaTech (UPC), ESEIAAT, Terrassa, Spain;
jose.gibergans@upc.edu ; carme.hervada@upc.edu

[2] U. Politècnica de Catalunya - BarcelonaTech (UPC), ETSECCPB, Barcelona, Spain;
eusebi.jarauta@upc.edu

Summary

The 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 (Hervada-Sala et al., 2018). 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, they have as Plaia and Ruggeri (2011) pointed out. In general, all indices ignore the compositional nature of the concentrations of air pollutants and do not apply methods of Compositional Data Analysis; those indices also have some other weak points such as leak of standardized scale.

A first approach applying Compositional Data Analysis methods has been developed in Jarauta-Bragulat et al., 2016. In the present work, we try to go some step further to improve the understanding and manageability of air quality. The air quality index proposed here takes into account the compositional nature of the data, it 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).

Keywords: Air pollution, Air quality, Air quality index, Health impact, Compositional Data Analysis, Log-ratio.

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. 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 (Hawking, 2017). 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 µg/m3. 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. For more details, see Plaia and Ruggieri 2011 and Jarauta-Bragulat et al. 2016.

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 Jarauta-Bragulat et al. 2016. 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, keeping as an essential element 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.

2. Proposal of a new definition of Air Quality 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:

$$M(k) = \begin{pmatrix} t_k & c_1(t_k) & c_2(t_k) & \cdots & c_D(t_k) \end{pmatrix},\ k = 1,2,\ldots,N \tag{2.1}$$

being t_k the k-time period (day, week, month, ...), $k = 1,2,\ldots,N$, N the number of time periods and $c_i(t_k)$ the concentration of ith air pollutant at time t_k (units are usually μg/m^3). The most significant air pollutants for their impact on people health in urban surroundings are O3, NO2, PM10 and PM2.5 (Arden-Pope and Dockery 2006; Zhang et al. 2016; Van den Elshout et al. 2014). 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

$$M(k) = \begin{pmatrix} t_k & c_{O3}(t_k) & c_{NO2}(t_k) & c_{PM10}(t_k) & c_{PM2.5}(t_k) \end{pmatrix}. \tag{2.2}$$

A logcontrast (LC) is defined as a linear combination of logarithms of parts with coefficients adding up to zero (Aitchison 1986). If a logcontrast is computed considering air pollutants and the residual component, for any time t_k, k=1,2,...,N, put $\alpha_{res} = -1$, then taking into account properties of logarithmic function, the equation we have can be written as:

$$\begin{gathered} LC = \log\left(c_{O3}^{\ \alpha_{O3}} c_{NO2}^{\ \alpha_{NO2}} c_{PM10}^{\ \alpha_{PM10}} c_{PM2.5}^{\ \alpha_{PM2.5}} \right) - \log\left(c_{res}\right) \\ \alpha_{O3} + \alpha_{NO2} + \alpha_{PM10} + \alpha_{PM2.5} = 1. \end{gathered} \tag{2.3}$$

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 Eq(2.3) is almost constant (Jarauta-Bragulat et al., 2016). Therefore, air pollution can be expressed using the considered air pollutants with Eq(2.3) or its approximation

$$\begin{gathered} f\left(c_{O3}, c_{NO2}, c_{PM10}, c_{PM2.5}\right) = c_{O3}^{\ \alpha_{O3}} c_{NO2}^{\ \alpha_{NO2}} c_{PM10}^{\ \alpha_{PM10}} c_{PM2.5}^{\ \alpha_{PM2.5}} \\ \alpha_{O3} + \alpha_{NO2} + \alpha_{PM10} + \alpha_{PM2.5} = 1. \end{gathered} \tag{2.4}$$

From Eq(2.4), a function for an Air Pollution Index (API) calculation can be stated as:

$$\begin{gathered} API = K_{global} \left(c_{O3}^{\ \alpha_{O3}} c_{NO2}^{\ \alpha_{NO2}} c_{PM10}^{\ \alpha_{PM10}} c_{PM2.5}^{\ \alpha_{PM2.5}} \right)^{\beta} \\ K_{global} = \frac{100}{f(O3_{\max}, NO2_{\max}, PM10_{\max}, PM2.5_{\max})} \\ \alpha_{O3} + \alpha_{NO2} + \alpha_{PM10} + \alpha_{PM2.5} = 1. \end{gathered} \tag{2.5}$$

The exponent β 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. 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 (Zhang et al. 2016, "air quality now, 2018"). An example is shown in Table 1 and the corresponding value of K_{global} is 100/223.8.

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

$$AQI^* = 100 - API \tag{2.6}$$

As a complement to the information provided by the global AQI* values, individual functions for each air pollutant can be calculated as:

$$
\begin{aligned}
API_{O3} &= K_{O3}\left(c_{O3}{}^{\alpha_{O3}}\right)^{1.25}; \quad AQI^*_{O3} = 100 - API_{O3} \\
API_{NO2} &= K_{NO2}\left(c_{NO2}{}^{\alpha_{NO2}}\right)^{1.25}; \quad AQI^*_{NO2} = 100 - API_{NO2} \\
API_{PM10} &= K_{PM10}\left(c_{PM10}{}^{\alpha_{PM10}}\right)^{1.25}; \quad AQI^*_{PM10} = 100 - API_{PM10} \\
API_{PM2.5} &= K_{PM2.5}\left(c_{PM2.5}{}^{\alpha_{PM2.5}}\right)^{1.25}; \quad AQI^*_{PM2.5} = 100 - API_{PM2.5}
\end{aligned}
\tag{2.7}
$$

	O3	NO2	PM10	PM2.5	*f*	Global AQI*
Scaling factor	0.0	0.0	0.0	0.0	0.0	100
0.10	24.0	40.0	20.0	15.0	22.4	94.4
0.40	96.0	160.0	80.0	60.0	89.5	68.2
0.65	156.0	260.0	130.0	97.5	145.4	41.6
0.85	204.0	340.0	170.0	127.5	190.2	18.4
	240.0	**400.0**	**200.0**	**150.0**	**223.8**	**0.0**

α_i	0.25	0.20	0.30	0.25

Table 1. Breakpoints for each considered air pollutants; in bold, maximum values assigned. First column indicates the factor to obtain partial breakpoints from corresponding maximum. Column *f* indicates the result according to Eq(2.4) and exponents are at the last row. In last column, global AQI* values and corresponding colour codes.

3. Case study: Barcelona 2001-2015

A data series 2001-2015 of air pollution in Barcelona is studied. In Figure 1, AQI* functions are plotted for the considered air pollutants. In Figure 2 global AQI* function is plotted. For this period, average of global AQI* values is 90.95, a reasonably good value for air quality. Average individual AQI* values are 91.36 for O3, 93.80 for NO2, 88.40 for PM10 and 86.22 for PM2.5.

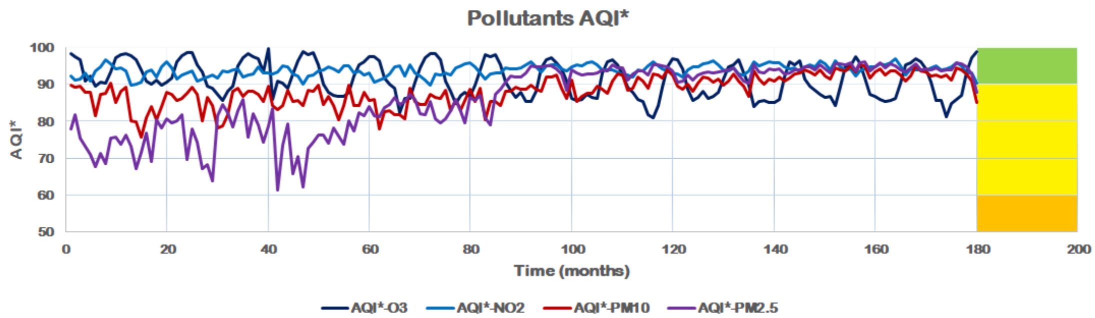

Figure 1. Individual AQI* functions for air pollutants.
Data series: Barcelona 2001-2015.

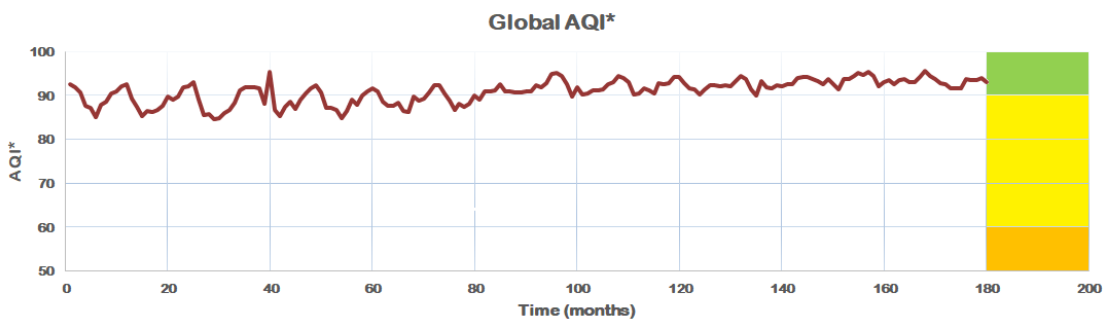

Figure 2. Global AQI* function.
Data series: Barcelona 2001-2015.

4. Conclusions

The main conclusions from this work are:
(1) Definition of an index of atmospheric air quality in cities should be formulated according to the compositional nature of the concentrations and, consequently, applying concepts and methods of Compositional Data Analysis.
(2) The function that should be used in the formulation of an index of air quality is the logcontrast.
(3) The methodology presented in this work allows characterizing air quality individually for each air pollutant, obtaining a global quality index and presenting an AQI-report that illustrates the evolution of global air quality.
(4) Attending to air quality in Barcelona, 2001-2015: (a) Global air quality is satisfactory and has improved in the last ten years. (b) The impact on air quality is mainly due to the presence of particles rather than the presence of gases. (c) In relation to particles, a significant increasing in both corresponding individual quality in last ten years can be

observed. (d) The individual index of ozone has worsened in the last ten years.

Acknowledgements

The authors are grateful to:
(1) The Spanish Ministry of Economy and Competitiveness (MINECO) for the financial support of the research project Fertilecity II "Integrated rooftop greenhouses: energy, waste and CO2 symbiosis with the building. Towards foods security in a circular economy" (CTM2016-75772-C3-1-R; CTM2016-75772-C3-2-R);
(2) The Spanish Ministry of Economy and Competitiveness (MINECO), within the framework of the "CODA-RETOS / TRANSCODA" Project (Ref. MTM2015-65016-C2-1-R and MTM2015-65016-C2-2-R);
(3) The Agency for Management of University Grants and Research (AGAUR) of the Generalitat de Catalunya (GENCAT) within the framework of the project "Analysis of spatial and compositional data" (COSDA, Ref: 2014SGR551, 2014-2016).

References

Aitchison J (1986). The Statistical Analysis of Compositional Data. Monographs on Statistics and Applied Probability. Chapman & Hall Ltd. London (UK).

Arden-Pope C, Dockery DC (2006). Health Effects of Fine Particulate Air Pollution: Lines that Connect. Journal of Air & Waste Management Assoc.56:709-742

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

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

Jarauta-Bragulat E, Hervada-Sala C, Egozcue JJ (2016). Air Quality Index Revisited from a Compositional Point of View. Mathematical Geosciences, 48:581-593.

Plaia A, Ruggieri M (2011). Air quality indices: a review. Rev Environ SciBiotechnol 10:165–179.

Van den Elshout S, Léger K, Heich H (2014). CAQI Common Air Quality Index - Update with PM2.5 and sensitivity analysis. Science of the Total Environment 488–489:461–8. doi.org/10.1016/j.scitotenv.2013.10.060.

Zhang J, Zhang LY, Du M, ZhangW, Huang X, Zhang YQ,Yang YY, Zhang JM, Deng SH, Shen F, Li YW, Xiao H (2016). Identifying the major air pollutants base on factor and cluster analysis, a case study in 74 Chinese cities. Atmospheric Environment 144:37-46.