Differential mortality risks associated to PM2.5 components: a multi-country multi-city study

Supplemental Digital Content

A. Data collection

A.1. Mortality

We obtained mortality data from the Multi-City Multi-Country (MCC) database. The current analysis was limited to cities that have air pollution, temperature, and urban characteristics indicator data in the same time frame as the $PM_{2.5}$ composition dataset. It includes a total of 210 urban areas in 16 countries/regions: Australia (3 cities, 2000–2009), Canada (19 cities, 1999-2015), Chile (4 cities, 2008–2014), China (3 cities, 2013–2015), Estonia (1 city, 2008–2015), Finland (1 city, 1999–2014), Germany (11 cities, 2004-2015), Greece (1 city, 2007–2010), Japan (36 cities, 2011–2015), Mexico (3 cities, 2003–2012), Portugal (1 city, 2004–2017), Spain (15 cities, 2004–2014), Sweden (1 county, 2001–2010), Switzerland (4 cities, 1999–2013), United Kingdom (25 cities, 1999–2016), and United States (82 cities, 1999–2006).

In the present study, mortality is represented by daily counts of either non-external causes (International Classification of Diseases, ICD-9: 0-799; ICD-10: A00-R99) or, where not available, all-cause only. Countries/regions with mortality from non-external causes include: Australia, China and Spain. Countries/regions with mortality from total causes include: Canada, Chile, Estonia, Finland, Germany, Greece, Japan, Mexico, Portugal, Switzerland, Sweden, United Kingdom, and United States.

A.2. Exposure

We obtained daily 24-h average concentrations of $PM_{2.5}$ in 210 cities. The geographic distributions of cities with $PM_{2.5}$ data and the corresponding annual-mean concentrations during respective study periods are shown in Figure 1 (main manuscript). We also collected daily mean temperature for the 210 cities in the analysis. In brief, measurements for air pollutants were obtained from fixed site monitoring networks operated by local authorities. The majority of monitors were located in urban areas, and only those daily measurements reporting above 75% of hourly data were included. On average, there were 4.7 monitors per city (ranging from 1 to 28), and measurements were averaged among all available monitors within one city to represent the exposure levels of the general population. In the main statistical analyses, we excluded the highest 5% and lowest 5% of $PM_{2.5}$ measurements to avoid the potential consequences in relation to inaccuracies driven by the outlying data points.

A.3. $PM_{2.5}$ composition

We extracted annual concentration of sulfate (SO_4^{2-}), nitrate (NO_3^-), ammonium (NH_4^+), black carbon (BC), organic carbon (OC), mineral dust (DUST) and sea salt (SS) for the 210 selected MCC cities from the V4.GL.03 dataset produced by the Dalhousie University Atmospheric Composition Analysis Group and freely available at http://fizz.phys.dal.ca/~atmos/martin/?page_id=140. This dataset is produced at a 1 x 1 km grid across the world by integrating information from several satellite products of aerosol optical depth (AOD) and the GEOS-CHEM chemical transport model, corrected by ground measurement. Validation of the dataset in North America indicates good agreement between reconstructed data and monitor data, with cross-validated R^2 between 0.57 and 0.96 (van Donkelaar et al., 2019). As the methodology is constant across regions with only monitor data for validation changing, reconstruction performances are expected to be similar for the rest of the dataset.

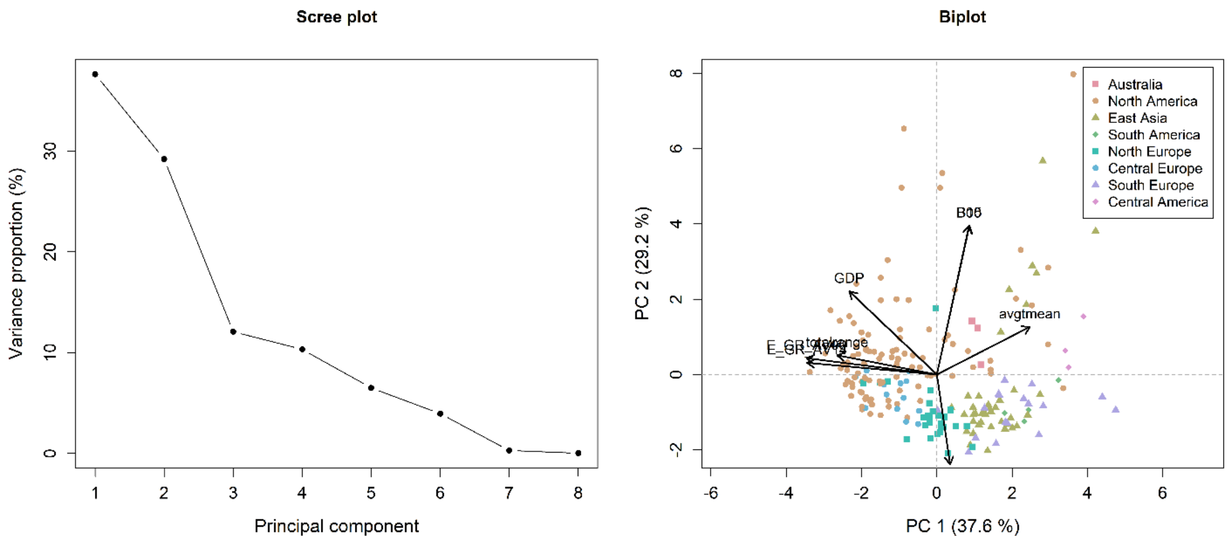
To link the components to MCC cities, all grid points of the composition dataset falling inside a circle of 10 km radius around the city reference location were aggregated. Obtained annual concentrations were then divided by the sum of all components for each year and location to obtain proportion of components. Note that we divided by the sum of components rather than the measured total $PM_{2.5}$ mass to account for the small reconstruction error in the sum, as the provided classification is supposed to be comprehensive. The annual compositions were then averaged using compositional data analysis tools to obtain single composition representing the whole period.

B. Socio-economic and environmental indicators PCA

eTable 1 details the socio-economic and environmental city-specific indicators considered in the analysis. The Scree plot of eFigure 1 shows that the two first components include 67% of the indicators' variance. According to the biplot, the first principal components (PC) represents environmental indicators by opposing average city greenness (E_GR_AV00 and E_GR_AV14) as well as temperature range (totalrange) to average temperature of the city (avgtmean). The second component is mainly driven by Built-up area proportion (B00 and B15) but also represents the proportion of people 65 years and older (oldpopprop). Finally, note that GDP is well represented by these two components.

eTable 1: Description of socio-economic and environmental indicators used in the PCA.

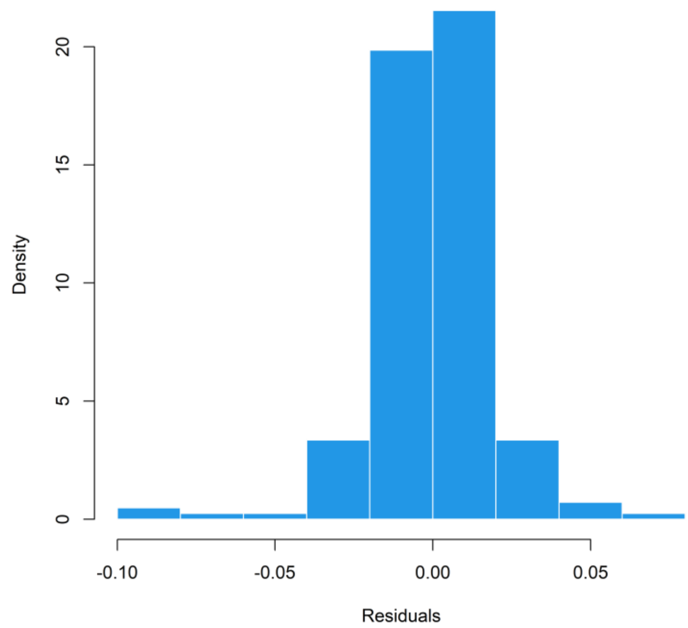
Name	Description	Time frame	Source
oldpopprop	Proportion of people age 65 years and above	2000	OECD Regional and Metropolitan Database
GDP	Gross domestic product per capita (in USD)	Mean 2001-2010	
avgtmean	Average temperature	City specific availability	MCC study database
totalrange	Total temperature range		
E_GR_AV00	Average greenness estimated through NDVI	2000	GHS Urban Center Database
E_GR_AV14		2014	
B00	Total Built-up area	2000	
B15		2015	



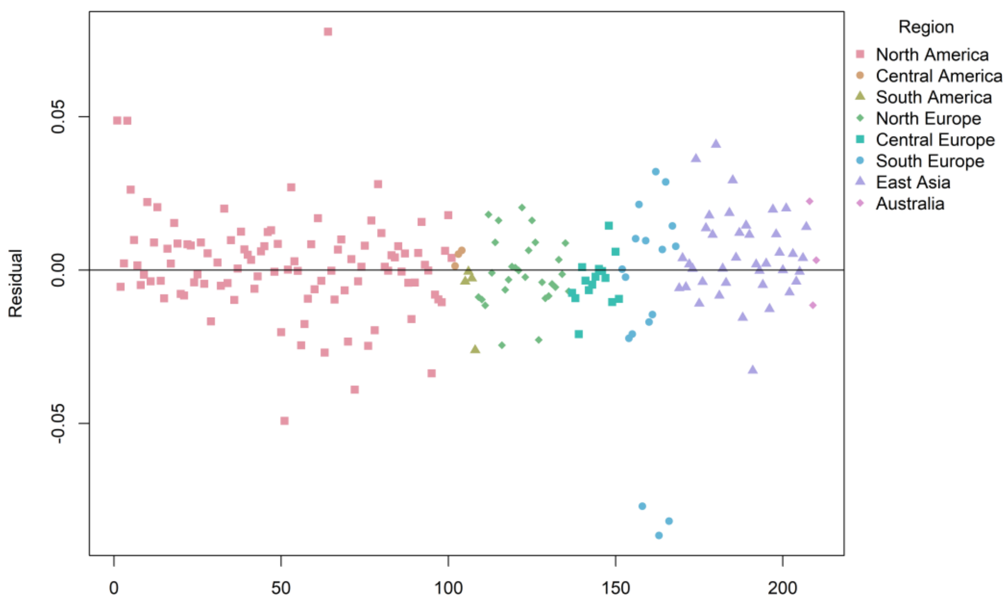
eFigure 1: Principal component analysis of socio-economic and environmental indicators. Scree plot showing the proportion of variance captured by each component (left) and biplot showing cities scores and indicators representation on the first two principal components.

C. Meta-regression residuals

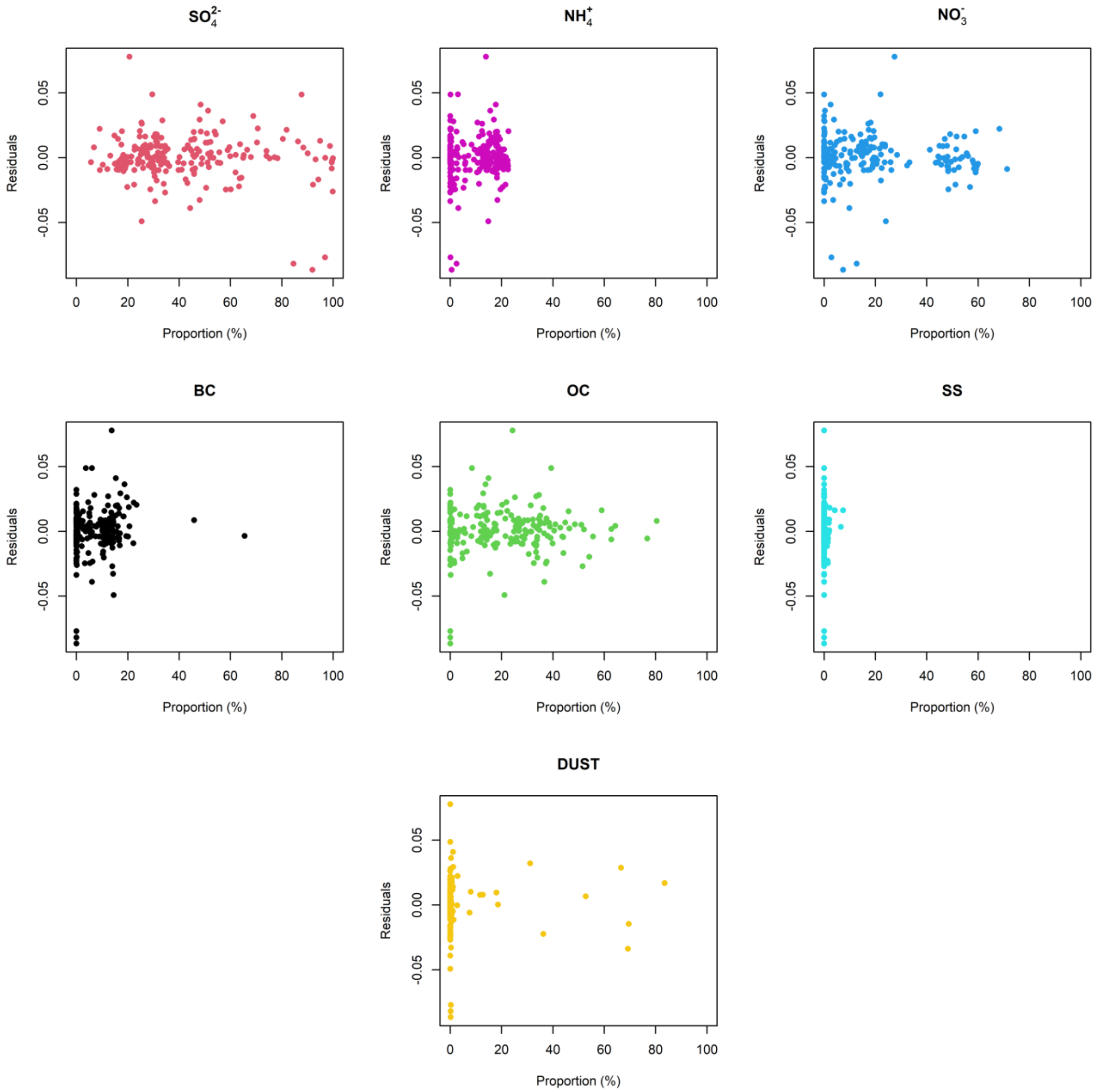
The histogram of eFigure 2 indicates that, except for few outliers with negative residuals (three Spanish cities as outlined in the main text), the residuals are roughly normal. In addition, according to eFigures 3 and 4, residuals do not indicate consistent bias, nor obvious heteroscedasticity. Therefore, the basic assumptions of a linear regression such as the second-stage meta-model are appropriate.



eFigure 2: Distribution of the residuals.

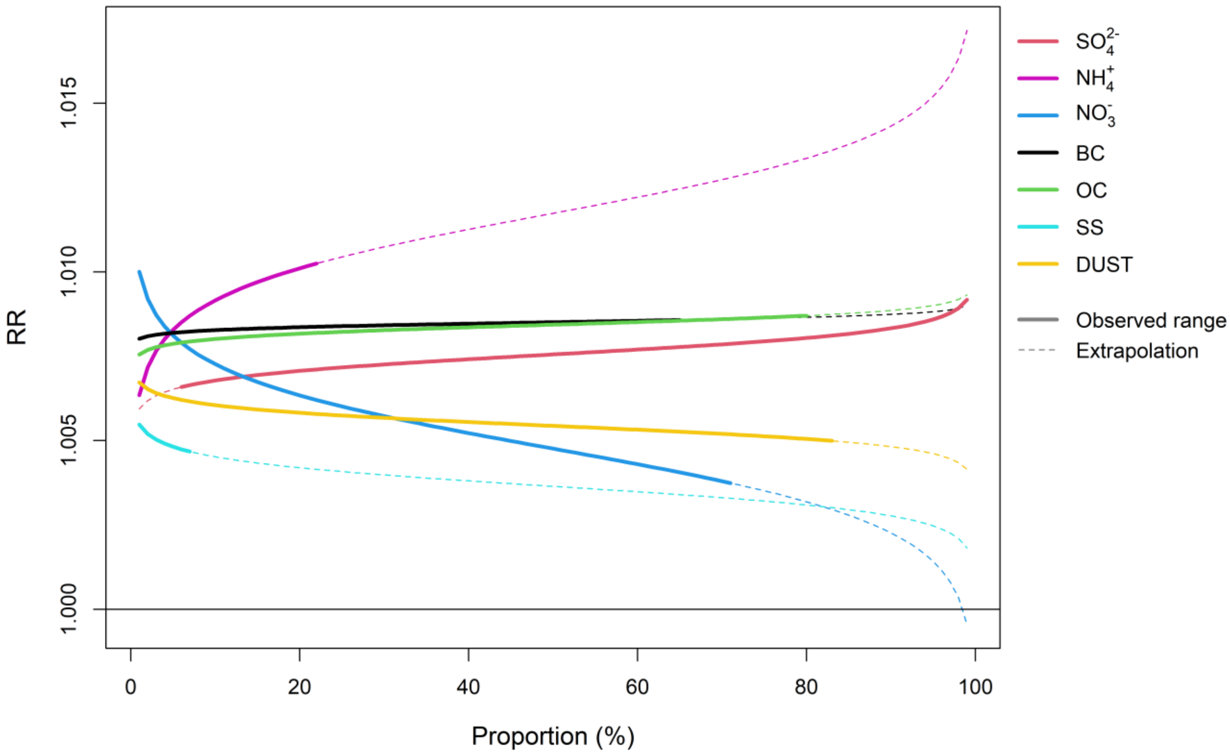


eFigure 3: Residuals of the meta-regression model clustered by the region of the city.

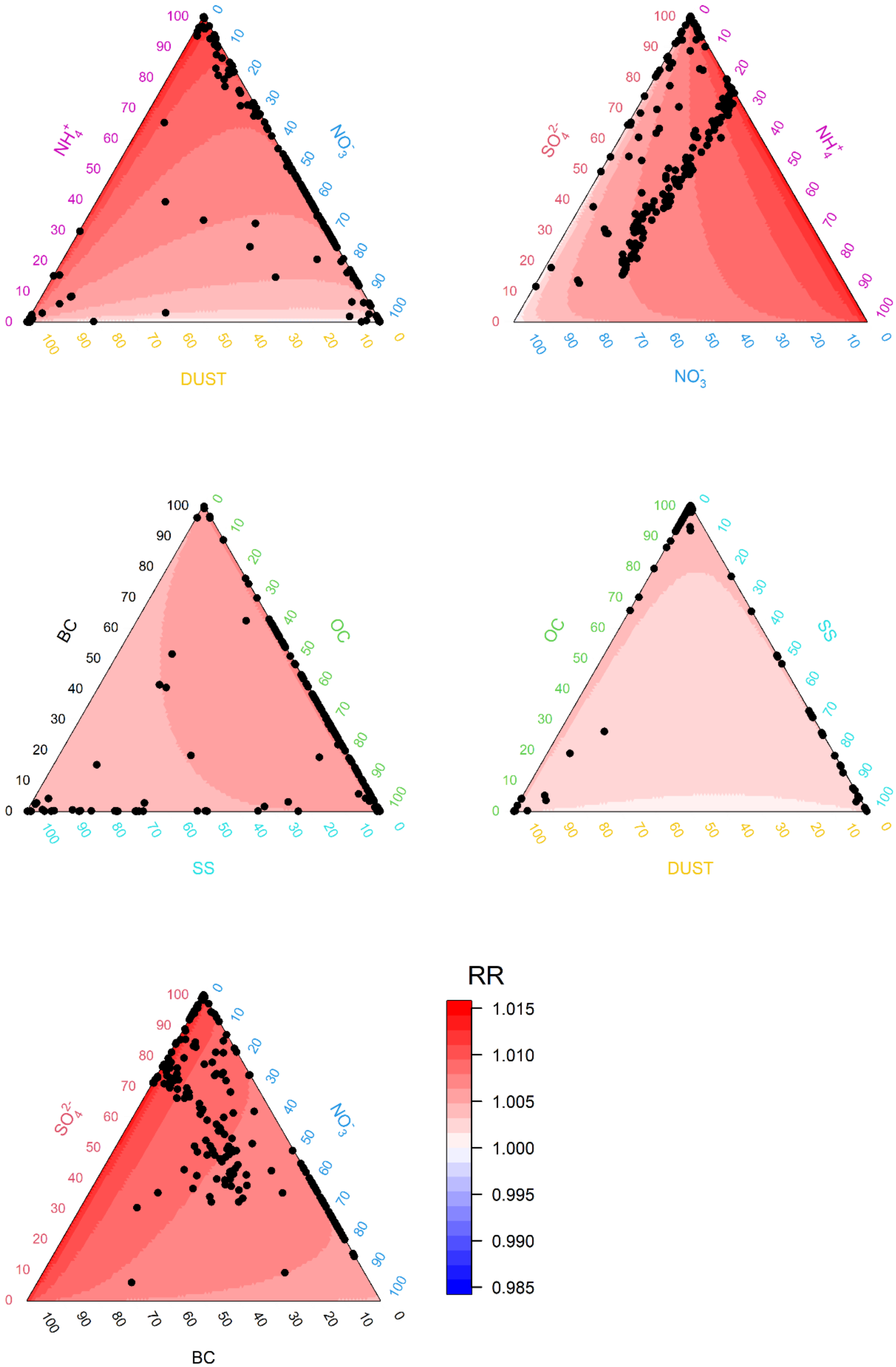


eFigure 4: Residuals of the meta-regression model versus the proportion of each component.

D. Alternative representation of results



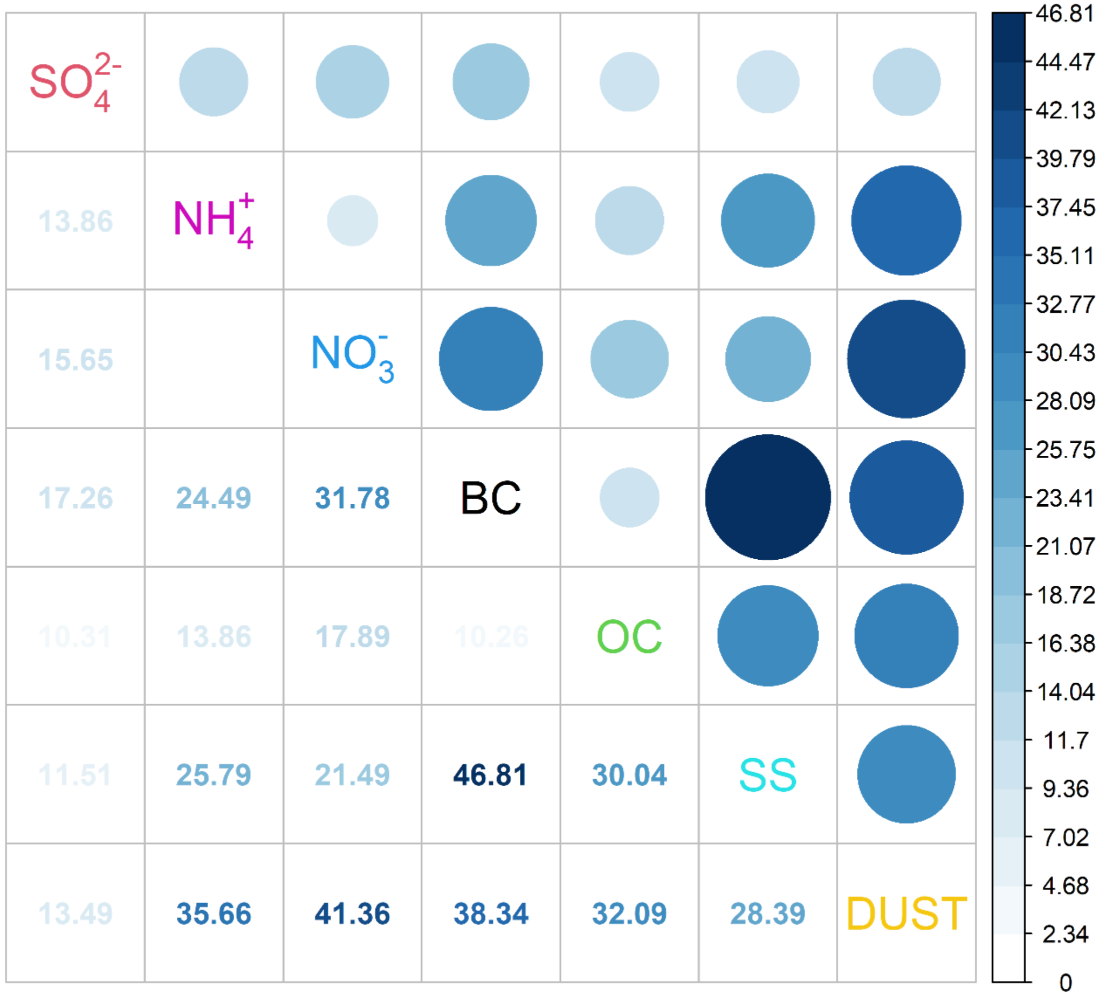
eFigure 5: Predicted relative risks (RRs) for different proportions of the components while keeping the other sub-composition constant. The predicted RR is associated to an increase of 10µg/m^3 of $PM_{2.5}$. Thick lines indicate the range of observed values for each component, while thin dashed lines indicate extrapolations.



NH_4^+
NO_3^-
DUST
SO_4^{2-}
NH_4^+
NO_3^-
BC
OC
SS
OC
SS
DUST
SO_4^{2-}
NO_3^-
BC
RR
1.015
1.010
1.005
1.000
0.995
0.990
0.985

eFigure 6: Ternary plots of predicted RR for different sub-compositions. Black dots represent the observed values.

E. Correlation between components and total $PM_{2.5}$



eFigure 7: Variation matrix of the $PM_{2.5}$ composition. The upper side colour and circle size represent the values displayed on the lower side of the diagonal. Variation between components x_j and x_k is defined as $var(log(x_j/x_k))$. A large variation value indicates that components tend to vary against each other. For instance, the large variation values associated to DUST mean that when it is present, other components tends to represent low proportions.

eTable 2: Correlation between components and total $PM_{2.5}$. The first line display the correlation between their absolute value and the total (the sum of components). The second line displays the correlation between the relative proportion and mean $PM_{2.5}$ computed in each city.

	SO4	NH4	NO3	BC	OC	SS	DUST
Absolute	0.77	**0.90**	0.59	0.72	0.78	-0.04	0.33
Relative	0.03	**0.13**	-0.04	0.01	-0.02	-0.08	-0.07