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

Pierre Masselot[1*], Francesco Sera[1,2], Rochelle Schneider dos Santos[1,3,4], Haidong Kan[5], Éric Lavigne[6,7], Massimo Stafoggia[8], Aurelio Tobias[9], Hong Chen[10], Richard T. Burnett[10], Joel Schwartz[11], Antonella Zanobetti[11], Michelle Bell[12], Bing-Yu Chen[13], Yue-Liang Leon Guo[13], Martina S. Ragettli[14], Ana Maria Vicedo-Cabrera[1,15,16], Christofer Åström[17], Bertil Forsberg[17], Carmen Íñiguez[18,19], Rebecca M. Garland[20,21,22], Noah Scovronik[23], João Paulo Teixeira[24], Baltazar Nunes[25, 26], César De la Cruz Valencia[27], Magali Hurtado Diaz[27], Yasushi Honda[28,29], Masahiro Hashizume[30], Evangelia Samoli[31], Klea Katsouyanni[31,32], Alexandra Schneider[33], Veronika Huber[34,35], Niilo Ryti[36,37], Jouni J.K. Jaakkola[36,37], Marek Maasikmets[38], Hans Orru[39], Yuming Guo[40], Nicolás Valdés Ortega[41], Patricia Matus Correa[41], Shilu Tong[42,43,44], Antonio Gasparrini[1,3,45]

1 Department of Public Health, Environments and Society, London School of Hygiene and Tropical Medicine (LSHTM), 15-17 Tavistock Place, London, WC1H 9SH, UK
2 Department of Statistics, Computer Science and Applications "G. Parenti", University of Florence, Florence, Italy
3 Centre on Climate Change and Planetary Health, London School of Hygiene & Tropical Medicine (LSHTM), Keppel Street, London, WC1E 7HT, UK
4 European Centre for Medium-Range Weather Forecast, Reading, UK
5 Department of Environmental Health, School of Public Health, Fudan University, Shanghai, China
6 School of Epidemiology and Public Health, Faculty of Medicine, University of Ottawa, Ottawa, Canada
7 Air Health Science Division, Health Canada, Ottawa, Canada
8 Department of Epidemiology, Lazio Regional Health Service/ASL Roma 1, Rome, Italy
9 Institute of Environmental Assessment and Water Research, Spanish Council for Scientific Research, Barcelona, Spain
10 Health Canada, Ottawa, Canada
11 Department of Environmental Health, Harvard T.H. Chan School of Public Health, Boston, MA, USA
12 School of the Environment, Yale University, New Haven CT, USA
13 National Institute of Environmental Health Science, National Health Research Institutes, Zhunan, Taiwan
14 Swiss Tropical and Public Health Institute, Basel, Switzerland
15 Institute of Social and Preventive Medicine, University of Bern, Bern, Switzerland
16 Oeschger Center for Climate Change Research, University of Bern, Bern, Switzerland
17 Department of Public Health and Clinical Medicine, Umeå University, Sweden
18 Department of Statistics and Computational Research. Universitat de València, València, Spain
19 Ciberesp, Madrid. Spain
20 Natural Resources and the Environment Unit, Council for Scientific and Industrial Research, Pretoria 0001, South Africa
21 Unit for Environmental Sciences and Management, North-West University, Potchefstroom 2520, South Africa
22 Department of Geography, Geo-informatics and Meteorology, University of Pretoria, Pretoria 0001, South Africa
23 Department of Environmental Health. Rollins School of Public Health, Emory University, Atlanta, USA
24 Department of Environmental Health, Instituto Nacional de Saúde Dr Ricardo Jorge
25 Department of Epidemiology, Instituto Nacional de Saúde Dr Ricardo Jorge
26 Centro de Investigação em Saúde Pública, Escola Nacional de Saúde Pública, Universidade NOVA de Lisboa
27 Department of Environmental Health, National Institute of Public Health, Cuernavaca, Morelos, Mexico
28 Center for Climate Change Adaptation, National Institute for Environmental Studies, Tsukuba, Japan
29 Faculty of Health and Sport Sciences, University of Tsukuba, Tsukuba, Japan
30 Department of Global Health Policy, Graduate School of Medicine, The University of Tokyo, Tokyo, Japan
31 Department of Hygiene, Epidemiology and Medical Statistics, National and Kapodistrian University of Athens, Greece
32 School of Population Health and Environmental Sciences, King's College, London, UK
33 Institute of Epidemiology, Helmholtz Zentrum München – German Research Center for Environmental Health (GmbH), Neuherberg, Germany

34 Potsdam Institute for Climate Impact Research, Potsdam, Germany
35 Department of Physical, Chemical and Natural Systems, Universidad Pablo de Olavide, Sevilla, Spain
36 Center for Environmental and Respiratory Health Research (CERH), University of Oulu, Oulu, Finland
37 Medical Research Center Oulu (MRC Oulu), Oulu University Hospital and University of Oulu, Oulu, Finland
38 Estonian Environmental Research Centre, Tallinn, Estonia
39 Department of Family Medicine and Public Health, University of Tartu, Tartu, Estonia
40 Department of Epidemiology and Preventive Medicine, School of Public Health and Preventive Medicine, Monash University, Melbourne, Australia
41 Department of Public Health, Universidad de los Andes, Santiago, Chile
42 School of Public Health and Social Work, Queensland University of Technology, Brisbane, Australia
43 School of Public Health and Institute of Environment and Human Health, Anhui Medical University, Hefei, China
44 Shanghai Children's Medical Centre, Shanghai Jiao-Tong University, Shanghai, China
45 Centre for Statistical Methodology, London School of Hygiene & Tropical Medicine (LSHTM), Keppel Street, London, WC1E 7HT, UK
**Corresponding Author: pierre.masselot@lshtm.ac.uk*

Abstract

Background. The association between $PM_{2.5}$ and mortality widely differs from country-to-country as well as within countries. Differences in $PM_{2.5}$ composition can play a role in determining differential risks, but there is little evidence about which components have larger impacts on mortality.

Objectives. To assess the role of the $PM_{2.5}$ composition on its associated risk and identify potentially harmful components through the statistical framework of compositional data analysis.

Methods. We applied a two-stage analysis on data collected from 202 locations in 18 countries. In the first stage, a relative risk for mortality associated with $PM_{2.5}$ was estimated for each city through a time series regression analysis. The estimates were then pooled in a second-stage meta-regression model that included city-specific average $PM_{2.5}$ composition as well as meta-predictors derived from socio-economic and large-scale environmental indicators. The $PM_{2.5}$ components were represented by sulfate (SO_4^{2-}), nitrate (NO_3^-), ammonium (NH_4^+), black carbon (BC), organic carbon (OC), mineral dust (DUST), and sea salt (SS). They were included in the meta-regression model through an additive log-ratio transformation to enforce a sum-to-one constraint.

Results. We found strong evidence that mortality risk varies depending on the proportion of some $PM_{2.5}$ components. Specifically, an increase of relative levels of NH_4^+ from 0 to 20% was associated with a RR of $PM_{2.5}$ on mortality increase from 1.005 to 1.009. Conversely, locations with higher levels of NO_3^- or DUST presented RR decrease from 1.008 to 1.004 and from 1.004 to 1.000 at their highest proportion respectively. No change in risk was found for variations of the

proportion of BC, OC, and SS. Differences in composition explained a substantial part of the heterogeneity in $PM_{2.5}$ risk.

Discussion. This study indicates that mortality risks associated with $PM_{2.5}$ are enhanced by a higher proportion of ammonium in the composition of the particulate, while the risk decreases in the presence of large concentrations of nitrate and dust. These findings can contribute to identify more dangerous emission sources and to implement more effective policies to prevent health risks related to air pollution.

Introduction

Particulate matter is a major environmental risk factor to which the Global Burden of Diseases attributes between 4.1 and 5 million deaths in 2017 (Stanaway et al. 2018). In particular, the short-term impact of fine particulate matter ($PM_{2.5}$) on mortality has been well-studied and it is now firmly established (Atkinson et al. 2014; Rückerl et al. 2011). However, some heterogeneity is observed on the health impacts of air pollution, both between (Liu et al. 2019) and within countries (Chen et al. 2017; Franklin et al. 2007).

A potential factor explaining such differences in health risks across populations is the variation in the chemical composition of $PM_{2.5}$. Particulate matter is a complex chemical mixture of various liquid or solid components varying in size, chemical composition, and other factors (Adams et al. 2015; Kelly and Fussell 2012). Some components are naturally present in the atmosphere and others emanate from anthropogenic activities, either as primary emissions or after chemical reactions in the atmosphere. The proportions of the components wildly vary across locations (McDuffie et al. 2020), and some may be more harmful than others.

Among the $PM_{2.5}$ components, previous studies have focused on black carbon/elemental carbon (BC/EC, thereafter only called BC), with systematic reviews suggesting a more important risk on all-cause mortality associated with BC alone when compared to the whole $PM_{2.5}$ concentration, both for short and long-term exposure (Janssen et al. 2011; Li et al. 2016). However, the review of Luben and colleagues (2017) found no particular impact on cardiovascular diseases for BC compared to $PM_{2.5}$. Sulfate (SO_4^{2-}) has also emerged as a potentially harmful component both in long-term cohort studies (Kioumourtzoglou et al. 2015; Ostro et al. 2010) and as a $PM_{2.5}$ short-term effect modifier (Franklin et al. 2008). The larger group of inorganic secondary aerosols (including SO_4^{2-}) has also been found as harmful in a recent cohort study in Denmark (Hvidtfeldt et al. 2019). Large scale studies and meta-analyses also suggest effects from specific metallic components such as nickel and vanadium, especially on cardiovascular and respiratory mortality (Bell et al. 2009; Yang et al. 2019). Nonetheless, the wide ranges of components and methodologies considered in these studies yield largely inconsistent results, in part due to studies being conducted in single locations or countries and in part by focusing on single components models (Achilleos et al. 2017). Explaining these differential risks is critical for developing and implementing effective actions to reduce health burdens related to air pollution.

A rigorous analysis of these associations requires disentangling the contributions of various $PM_{2.5}$ elements, a step that poses important methodological challenges. Individual components are highly correlated to each other, as well as to total $PM_{2.5}$, and components-specific estimates from separate models are likely to be affected by confounding from other components. Controlling for these biases is no simple matter, mainly due to the nature of such data, i.e. the sum-to-one constraint of the composition (Butler 1979; Mostofsky et al. 2012). To address statistical issues posed by such constraint, Aitchison (1986) and references therein developed the coherent and elegant theory of *compositional data analysis*. This theory led to the adoption of *logratio transforms*, which can then be used in standard statistical methods, including regression (Aitchison and Bacon-Shone 1984; Hron et al. 2012). However, this methodology has been rarely used in epidemiological analyses on the health effects of $PM_{2.5}$ components. To the best of our knowledge, the sole exception is the study by Crouse and colleagues (2016), although without specifically referring to the statistical theory of Aitchison.

The objective of the present study is to identify and compare the all-cause mortality risks associated with constituents of $PM_{2.5}$ through the application of compositional data analysis methods, using a large international dataset gathered within the Multi-Country Multi-City Collaborative Research Network (MCC).

Methods

Data

Data include daily time series of all-cause mortality, $PM_{2.5}$ concentration, and temperature, as well as annual $PM_{2.5}$ composition and socio-economic indicators for 202 locations (exclusively urban areas) belonging to 18 countries included in the MCC dataset. This dataset, including the derivation of city-specific $PM_{2.5}$ series, is well described in Liu et al. (2019). Cities included in the present study must have at least one common year of record for each of the variables used in the study. Daily time series for the selected cities include record lengths spanning one to 18 years, with the earliest being 1999 and the latest 2017 to roughly match the availability of composition data. We assume the association between $PM_{2.5}$ and mortality did not significantly change in the last 20 years, allowing us to extend first-stage time-series length compared to the availability of composition data, and thus obtain more accurate RR estimates. This assumption has been checked with graphical tools. Table 1 provides details about data for each represented country.

To control for confounding from location-specific socio-economic and environmental indicators potentially correlated with specific composition patterns, we collected the proportion of people aged 65 and above in 2000, the average of gross domestic product per capita (GPD) between 2001 and 2010, average poverty rate after taxes and transfers between 2009 and 2014, as well as from the Organisation for Economic Co-operation and Development (OECD) Regional and Metropolitan Database (Maraut et al. 2008; Sera et al. 2019b). In addition, we consider total built-up area in 2000 and 2015, and average greenness estimated for 2000 and 2014, gathered from the GHS Urban Centre Database (Florczyk et al. 2019). Details are given in Supplemental Material A (Table S1).

$PM_{2.5}$ composition

We extracted $PM_{2.5}$ composition estimates for all MCC cities from the Dalhousie University Atmospheric Composition Analysis Group website (http://fizz.phys.dal.ca/~atmos/martin/?page_id=140). These estimates are available annually between 2003 and 2017 on a grid of 1km by 1km. To attribute a value to a city, we extract and average all grid points on a buffer of 10 km around the city reference location.

Estimates of $PM_{2.5}$ concentration are obtained via multiple satellite-based retrievals of aerosol optical depth in combination with the GEOS-Chem Chemical Transport Model, and enhanced through statistical incorporation of ground-based observations, as described in van Donkelaar et al. (2019). This yields partitioned data of seven components that are sulfate (SO_4^{2-}), nitrate (NO_3^-), ammonium (NH_4^+), the three of them forming the group of secondary inorganic aerosols, as well as black carbon (BC), organic carbon (OC), mineral dust (DUST) and sea salt (SS). These components provide a comprehensive classification of the main sources of $PM_{2.5}$.

SO_4^{2-} and NO_3^- are secondary inorganic components that originate from the oxidation of sulphur and nitrogen oxides, whose sources include fossil fuel combustion (gas and oil) as well as volcanoes. The third secondary inorganic aerosol, NH_4^+, originates mainly from fertilizer use and livestock (Park et al. 2004). Organic components, OC and BC, are emitted by all types of combustion, being more associated with residential sources such as biofuel than NO_3^- (McDuffie et al. 2020). In some countries such as Canada and Australia, OC is also associated with wildfires

(Meng et al. 2019). BC (sometimes called black smoke or elemental carbon) is more related to transportation (Bond et al. 2004). DUST mainly contains coarser particles transported from deserts (Hashizume et al. 2020; Stafoggia et al. 2016), but it can also include industrially emitted particles such as metals and cement (Philip et al. 2017). Finally, SS originates from sea spray and is thus more prominent in coastal areas (van Donkelaar et al. 2019).

Statistical analysis

The statistical analysis follows a two-stage design, first estimating a relative risk (RR) for $PM_{2.5}$ at the city level, and then modeling the heterogeneity of these RR in a meta-regression model. The analysis is entirely performed using the R software version 4.0.3 (R Core Team 2020) with additional packages `dlnm` (Gasparrini 2011), `mixmeta` (Sera et al. 2019a), `compositions` (van den Boogaart and Tolosana-Delgado 2008), and `zCompositions` (Palarea-Albaladejo and Martín-Fernández 2015).

First-stage modeling

At the city level, we performed a time series analysis with a quasi-Poisson regression model, following the specification of a previously published study (Liu et al. 2019). Briefly, $PM_{2.5}$ entered the model linearly as a 2-day moving average to account for both concurrent and lag-1 delayed effects. We accounted for confounding by temperature by including a natural spline of its 4-day moving average with knots at the 10^{th}, 75^{th}, and 90^{th} percentiles. Finally, the model also included a factor for day-of-week to account for weekly cycles in mortality and a natural spline of time with 7 degrees of freedom per year to account for seasonal effects.

Second-stage modeling

The second stage consists of a two-level random-effects meta-regression (Sera et al. 2019a) using the $PM_{2.5}$ components as meta-predictors. Random effects are added at the city and country level, allowing to control for confounding due to structural differences at different grouping levels and spatial scales. Besides, we accounted for potential confounding from the long list of socio-economic and large-scale environmental variables described above by including in the meta-regression model their first two principal components (PC), which accounted for 58% of this dataset's variance (Figure S1 in Supplemental Material A). The inclusion of the $PM_{2.5}$ components to the meta-regression model follows a compositional data approach as described below. To interpret the RR at the city level, we report the best linear unbiased predictions (BLUPs) from the meta-regression model described above (Gasparrini et al. 2012). Finally, we also checked the residuals to ensure that there is no obvious bias, heteroscedasticity, or departure from normality (see Supplemental Material B).

We quantified heterogeneity between locations with standard measures of I^2 and Cochran Q (Higgins and Thompson 2002). To assess how much effect modification is brought by variation in the components, these measures were estimated first from a meta-analysis model without meta-predictors (the "null" model), then from a model with only the PC of confounding indicators, and finally from the full meta-analysis model that includes the components. In addition, a Wald test is performed between nested models to assess the reduction in residual heterogeneity provided by the $PM_{2.5}$ composition. All the models were fitted through restricted maximum likelihood (REML).

Compositional data and logratio transforms

The basic definition of a compositional dataset is a collection of variables $x_1,\ldots,x_D$ such that $x_j>0 \forall j$ and $\sum_{j=1}^{D} x_j=1$, where $D=7$ in the present analysis. Because of this sum-to-one constraint, the components x_j are necessarily correlated, and the compositional data only provides information about the relative variations of the components to each other. This led Aitchison to develop the logratio approach of compositional data analysis in a series of papers (Aitchison 1981, 1982, 1983), to consider quantities $\log\left(x_j/x_k\right)$. A logratio gives information about the relative proportion of the components x_j and x_k in a symmetric way.

The basic process of compositional data analysis is to transform the compositional dataset $x_1,\ldots,x_D$ into $D-1$ new variables through the *additive logratio* (ALR) transformation:

$$z_j=\log\left(\frac{x_j}{x_D}\right) \tag{1}$$

for $j=1,\ldots,D-1$, using the D^{th} component as the baseline comparison. This transformation allows removing the sum-to-one constraint while retaining the relative information of all components. Classical statistical analyses can then be performed on the z_j variables. Note that the final results are insensitive to the chosen baseline component x_D in equation (1) (Aitchison 1986, chapter 5).

As summary statistics of the $PM_{2.5}$ composition, we computed the compositional mean of each country. We thus transformed the data of each year and each city using the ALR of equation (1) and computed their mean by country. The compositional mean was then obtained by back-transforming the mean of z_j variables (Aitchison 1982). As a compositional equivalent to correlations, we also computed the variation matrix of the compositions. The variation matrix contains at position (j,k) the value $var(\log(x_j/x_k))$ which intuitively represents how much the two components vary relative to each other (Aitchison 1986). A large variation value would roughly mean that the components tend to replace each other.

For the second-stage meta-analysis, we first transformed the composition through ALR, we averaged these ALR by city, and we used the ALR transformed averages as meta-predictors. This results in the meta-regression model:

$$\log\left(RR_{ij}\right)=\beta_0+\sum_{k=1}^{D-1}\beta_j\log\left(\frac{x_{ijk}}{x_{ijD}}\right)+\gamma_1 PC_{ij1}+\gamma_2 PC_{ij2}+\omega_j+\xi_{ij}+\epsilon_{ij}$$

$$\text{¿}\,\beta_0+\sum_{j=1}^{D}\beta_j\log\left(x_{ijk}\right)+C_{ij} \tag{2}$$

where $\beta_D=-\sum_{j=1}^{D-1}\beta_j$. $\log\left(RR_{ij}\right)$ is the $PM_{2.5}$ coefficient obtained in the first stage of the analysis for city i of country j. In the second line of (2), the term $C_{ij}=\gamma_1 PC_{ij1}+\gamma_2 PC_{ij2}+\omega_j+\xi_{ij}+\epsilon_{ij}$ includes all location-level confounding, where PC_{ijm} indicate the indicators' principal components, ω_j and ξ_{ij} are respectively the country and city-level random effects and ϵ_{ij} is the residual. The second representation in (2) shows that the coefficients β_1 to β_D sum to zero because we deal with a composition, and that the final results do not depend on the baseline component x_D in the ALR. Using the ALR as in the first representation in (2) is useful for model fitting since it removes the sum-to-zero constraint of the coefficients.

To interpret the results of the meta-regression model in (2), we predict the expected RR for a range of values of each component x_j while keeping the sub-composition of other components constant under the sum-to-one constraint. For each value of x_j between 0 and 1, we set the other components equal to their mean over the whole dataset, adjusting it to keep their sum to 1. Thus, only the relative proportion of x_j to the other components varies, while the relative proportion of the other components between them stays constant. In addition, the two principal components of the socio-economic and environmental indicators are set to zero. The prediction obtained with the model in (2) is then exponentiated to report it on the RR scale.

The $PM_{2.5}$ composition dataset contains many zero values (especially SS and DUST) which are not allowed in the logratio analysis for obvious reasons. We thus consider the imputation procedure of Martín-Fernández et al. (2003) that replaces zeros by small values (here equal to 10^{-5}) and adjusts other components accordingly to keep the sum-to-one constraint as well as the relative values between components.

Results

Descriptive statistics

Table 1 shows some summary statistics of the mortality and pollution data aggregated per country. The data used for the first stage span around 10 years on average, scattered between 1999 and 2016. A total of more than 15 million deaths occurred overall. Figure 1 shows all the cities with mean observed $PM_{2.5}$. The highest levels of $PM_{2.5}$ are observed in China, Taiwan, South-Africa as well as South-America. On the other hand, the lowest $PM_{2.5}$ levels are observed in northern countries (i.e. Sweden and Canada) as well as in Australia.

Figure 2 shows the mean $PM_{2.5}$ composition in each country. Overall, no obvious pattern can be found in the composition, although some countries show widely variable distributions. The wider variability is observed in countries affected by DUST since it can represent a significant part of $PM_{2.5}$ one year and be almost absent the next one. Countries particularly impacted are Australia and Mediterranean countries (Greece, Portugal, and Spain). Overall, the two most important components seem to be SO_4^{2-} and NO_3^-, both linked to the burning of fossil fuel. NO_3^- is more represented in European countries except for Mediterranean ones, while SO_4^{2-} is widely present in hotter countries. OC represents an important part of the composition in Nordic countries since it is linked to wildfires but also residential wood burning. BC and NH_4^+ are overall lower parts of the $PM_{2.5}$ composition, and SS is only present in seaside countries, notably Portugal and the UK.

The variation matrix in Figure 3 shows how the components vary in relation to each other. In particular, DUST varies substantially against other components (except SO_4^{2-}), since it tends to be a major part of the composition when present. We also observe a large variation between BC and SS, although this can be an artefact due to SS being present in only two countries that happen to have low proportions of BC. BC and NO_3^- also vary against each other, which indicates the interaction between volatile organic components and secondary inorganic aerosols (Aksoyoglu et al. 2017).

Second stage city-specific relative risks

The BLUPs of RRs are reported in Figure 1 for each city and range from 0.995 in Valladolid (Spain) to 1.021 in Sendai (Japan). Predicted RRs are above 1 for 197 cities among the 202 used in the model. The highest RRs are found in North-America, Mexico, and Japan, as well as specific locations in Europe such as Switzerland and Greece. In contrast, lower predicted RRs are found in Europe.

Supplemental Material C provides insight on the location-specific residuals from the second-stage meta-regression. Their distribution indicates that the model overall fits well European cities, while residuals are slightly more variable for North-America and Asia. Six outliers can be seen on the residuals. Three Spanish cities RRs are overestimated by the model (San Sebastian, Vitoria, and Leon). Indeed, these cities correspond to low first-stage RRs despite elevated proportions of SO_4^{2-} (above 80%, see Figure S4). The three high residuals correspond to North American cities (Madison, Halifax, and Abbotsford) and which correspond to particularly high first-stage RRs despite average compositions.

Second-stage meta-regression interpretation

Predicted $PM_{2.5}$ RRs within observed ranges of each component are shown in Figure 4. A more direct comparison of the predicted curves, but without confidence intervals, along with ternary representations, are shown in Supplemental Material C. The logit form of reported curves stems from the ALR transformation applied to the components prior to the meta-regression model (see Equation 2). Results show a significantly positive effect modification of NH_4^+, suggesting that the effect of $PM_{2.5}$ is larger for cities with a higher relative proportion of NH_4^+ levels. RRs also increase with SO_4^{2-} although with a wider interval at lower proportions. Conversely, an increase in the proportion of NO_3^- and DUST results in a decrease of the RRs. Surprisingly, the RR curve is flat for carbonaceous components (BC and OC) meaning that the RR does not seem to be affected by relative variation in these components. Finally, although a slight decrease is seen, SS is too rare for inferring specific associations, as demonstrated by the wide confidence intervals. Note that three of the estimated coefficients used to obtain these predictions (the β_j of Equation (2)) differ substantially from the null value: NH_4^+ in a positive direction, while NO_3^- and DUST showing negative effect modification (Figure S7).

Table 2 shows that including the components as meta-predictors significantly reduces residual heterogeneity in the meta-analysis model. The Q statistics drops from 473 in the model fitted using only the socio-economic and environment PC model to 318 in the full model, with an I^2 of 58% and 39%, respectively. A Wald test on these nested models has a p-value of about 0.01, indicating that the composition explains a large part of the heterogeneity. Table 2 also shows that the model including only the socio-economic and environment PCs results in almost no added-value compared to the null model. This is confirmed by the fact that in the full model, the socio-economic and environment PCs are associated with approximately null coefficients. All these criteria concur in providing strong evidence that the heterogeneity in risk to $PM_{2.5}$ is in large part explained by its composition.

Discussion

The main finding of this study is the evidence of a role of ammonium (NH_4^+) in enhancing the mortality risks of $PM_{2.5}$. This is a component that has received little attention, compared to others such as BC, OC, and SO_4^{2-}, although it is one of the three secondary inorganic aerosols. Among the published studies reviewed by the authors (Bell et al. 2009; Chen et al. 2020; Franklin et al. 2008; Kioumourtzoglou et al. 2015; Peng et al. 2009; Wang et al. 2014), none of them reported a significant effect modification of NH_4^+. However, in a study on a Canadian cohort, Crouse et al. (2016) identified NH_4^+ as the component with the highest coefficient in a model that included all components and total $PM_{2.5}$ concentration. Note that this is the only study we are aware of that used a strategy similar to the compositional data approach considered in this contribution. Besides, few studies focusing on the concentration of components rather than their effect modification found positive associations between mortality and NH_4^+ (Huang et al. 2012; Lin et al. 2016; Liu and Zhang 2015; Son et al. 2012), although the analyses included many other components. In addition, confounding by total $PM_{2.5}$ concentration is rarely accounted for in these studies.

Interestingly, NH_4^+ shows low variation with the two other secondary inorganic components (see Figure 3), indicating that when NH_4^+ is high, it is likely that the whole proportion of secondary inorganic aerosols is important. Also, NH_4^+ is the most correlated component with the total $PM_{2.5}$ mass in our dataset (see Table S2). It has been suggested that ammonia, the main precursor of NH_4^+, is a major driver of $PM_{2.5}$, at least in some countries (Air Quality Expert Group 2013; Pinder et al. 2007; Wu et al. 2016). From a policy point of view, our study suggests that a larger focus on ammonia for mitigation strategies may provide important health benefits.

The other important result of our analysis is the observed reduction in RRs for high proportions of nitrate (NO_3^-) in the composition of $PM_{2.5}$. Indeed, NO_3^- represents a large part of the total concentration in northern European countries (Estonia, Finland, Germany, Switzerland, Sweden, and the UK, see Figure 2), which are areas displaying non-significant associations between $PM_{2.5}$ and mortality (Liu et al. 2019). NO_3^- is a secondary product of nitrogen oxides emissions, emitted by gas and oil burning, and is thus mainly related to traffic. Note that in the data used here, it presents an important compositional variation value with BC (see Figure 3), meaning that when it increases, NO_3^- tends to replace BC. Note that both are usually considered traffic-related components, NO_3^- being mainly related to oil and gas combustion while BC also includes all biofuel combustion (McDuffie et al. 2020). Therefore, related to traffic policies, mitigation strategies focusing more heavily on BC emissions compared to NO_3^- precursors may prove more effective from a public health point of view.

The model assessments suggest that the results reported above are robust to confounding by either socio-economic indicators or specific regional effects. Indeed, the socio-economic and environmental PCs added to the second-stage regression model contribute to explain a negligible part of the heterogeneity observed between cities (as reported in Table 2). The residual analysis does not show obvious patterns that may have been missed by the model either, since no regional or component-specific pattern emerges.

The regression results rely on the ALR proposed by Aitchison (1986) for compositional data analysis. However, note that other types of logratio transformation exist, such as centered logratio (CLR) and isometric logratio (ILR, Egozcue et al. 2003). CLR does not remove the closure constraint of compositional data and it is, therefore, difficult to use in a regression analysis. ILR enjoys the mathematical advantages of both ALR and CLR and it is popular (e.g. Mert et al. 2016), but it is less straightforward to use in regression analysis due to redundancy of information between the logratio variables. Hron and colleagues (2012) propose a procedure for regression with ILR that involves performing a separate regression model for each component, a procedure that gained popularity in recent years (Giancristofaro et al. 2020; Muller et al. 2018). However, it can be shown that this procedure is equivalent to the one used here, with identical coefficient estimates up to a scaling factor, while being less convenient than the approach considered here because of the requirement of multiple model fitting.

The strengths of the study lie in both the data used and the methods applied. It takes advantage of a large international dataset from the MCC network to evaluate how $PM_{2.5}$ composition affects its association with all-cause mortality. A wide heterogeneity in the composition is observed between locations, allowing the comparison of different compositional patterns. In addition, the study uses state-of-the-art statistical methods, including the recently proposed mixed-effect meta-analysis two-stage framework (Sera et al. 2019a) and compositional data analysis. The mixed-effect framework allows considering several levels of heterogeneity to the meta-analysis, which are country and city level here. This allows capturing heterogeneity at both levels, for instance related to differential country-wide and city-specific policies, as well as climatic or environmental conditions that may modify the association between $PM_{2.5}$ and all-cause mortality. Compositional data analysis provides a rigorous framework to analyse the role of different constituents of $PM_{2.5}$. Such data structures are prone to spurious results and misinterpretations if not analysed properly, as already observed by Pearson (1897). To the best of our knowledge, this study is the first to consider compositional data approaches to evaluate the effect modification of $PM_{2.5}$ composition.

Although the wide range of locations available is a strength of the study, it comes with the limitation that the measurement of total $PM_{2.5}$ differs across locations. A part of this uncertainty is nonetheless captured by the random effect on countries added to the model. However, the composition data we used are derived from remote sensing rather than station measurement. This provides a consistent measure of the compositions across locations. A side effect is that the sum of components does not always exactly add to the mean annual measured $PM_{2.5}$, due to complex interactions between diverse emission sources as well as uncertainties in the models generating the data. However, this difference is usually negligible (van Donkelaar et al. 2019).

The analysis performed here relies on the underlying assumption that the composition of $PM_{2.5}$ and its association with mortality have stayed roughly constant during the past 20 years. Figure 2 suggests that it is a reasonable assumption with few exceptions (UK and Greece). A potential extension of our approach would be to account for temporal differences, both as a long-term trend and as a seasonal pattern by using monthly data. However, this would require longer time-series than what is available for many countries in the MCC dataset, and it poses non-trivial methodological problems. This extension can be the topic of future research studies.

An important limitation related to compositional data analysis is the high number of zero values in the compositional dataset, especially for DUST and SS. It is recognized that the presence of zero values is an issue in compositional data analysis, and ad-hoc methods have been used to deal with it (Martín-Fernández et al. 2012).

The main message of the present paper is that $PM_{2.5}$ composition plays a significant role in the observed heterogeneity of mortality risk linked to air pollution and that it necessitates appropriate analytical methods. We hope that the present study will encourage researchers to make use of compositional data analysis tools in future studies. Surprisingly, we found that the most harmful component may be ammonium, while the widely studied black carbon and organic carbon components were found to have no impact on the health effects of $PM_{2.5}$. At the same time, a significant decrease in the health risk was found for higher proportions of nitrates. These results may suggest that specific action aimed at ammonia precursors, including the agricultural sector, as well as decreasing the part of BC compared to nitrates precursors in traffic-related emissions may prove effective in reducing the health impacts of air pollution.

Acknowledgements

We would like to thanks the Atmospheric Composition Analysis Group for providing the composition dataset used in the present study. This work was supported by the Medical Research Council of UK (Grant ID: MR/M022625/1), the Natural Environment Research Council of UK (Grant ID: NE/R009384/1), and the European Union’s Horizon 2020 Project Exhaustion (Grant ID: 820655).

Bibliography

Achilleos S, Kioumourtzoglou M-A, Wu C-D, Schwartz JD, Koutrakis P, Papatheodorou SI. 2017. Acute effects of fine particulate matter constituents on mortality: A systematic review and meta-regression analysis. Environ Int 109:89–100; doi:10.1016/j.envint.2017.09.010.

Adams K, Greenbaum DS, Shaikh R, Erp AM van, Russell AG. 2015. Particulate matter components, sources, and health: Systematic approaches to testing effects. J Air Waste Manag Assoc 65:544–558; doi:10.1080/10962247.2014.1001884.

Air Quality Expert Group. 2013. Mitigation of United Kingdom PM2.5 Concentrations. 49.

Aitchison J. 1981. A New Approach to Null Correlations of Proportions. Math Geol 13.

Aitchison J. 1983. Principal component analysis of compositional data. Biometrika 70:57–65; doi:10.1093/biomet/70.1.57.

Aitchison J. 1986. *The statistical analysis of compositional data*. Chapman & Hall, Ltd.:GBR.

Aitchison J. 1982. The Statistical Analysis of Compositional Data. J R Stat Soc Ser B Methodol 44:139–160; doi:10.1111/j.2517-6161.1982.tb01195.x.

Aitchison J, Bacon-Shone J. 1984. Log contrast models for experiments with mixtures. Biometrika 71:323–330; doi:10.1093/biomet/71.2.323.

Aksoyoglu S, Ciarelli G, El-Haddad I, Baltensperger U, Prévôt ASH. 2017. Secondary inorganic aerosols in Europe: sources and the significant influence of biogenic VOC emissions, especially on ammonium nitrate. Atmospheric Chem Phys 17:7757–7773; doi:https://doi.org/10.5194/acp-17-7757-2017.

Atkinson RW, Kang S, Anderson HR, Mills IC, Walton HA. 2014. Epidemiological time series studies of PM2.5 and daily mortality and hospital admissions: a systematic review and meta-analysis. Thorax 69:660–665; doi:10.1136/thoraxjnl-2013-204492.

Bell ML, Ebisu K, Peng RD, Samet JM, Dominici F. 2009. Hospital Admissions and Chemical Composition of Fine Particle Air Pollution. Am J Respir Crit Care Med 179:1115–1120; doi:10.1164/rccm.200808-1240OC.

Bond TC, Streets DG, Yarber KF, Nelson SM, Woo JH, Klimont Z. 2004. A technology-based global inventory of black and organic carbon emissions from combustion. J Geophys Res-Atmospheres 109:D14203; doi:10.1029/2003JD003697.

Butler JC. 1979. Effects of Closure on the Measures of Similarity Between Samples. Math Geol 11; doi:431-440.

Chen H, Zhang Z, van Donkelaar A, Bai L, Martin RV, Lavigne E, et al. 2020. Understanding the Joint Impacts of Fine Particulate Matter Concentration and Composition on the Incidence and Mortality of Cardiovascular Disease: A Component-Adjusted Approach. Environ Sci Technol 54:4388–4399; doi:10.1021/acs.est.9b06861.

Chen R, Yin P, Meng X, Liu C, Wang L, Xu X, et al. 2017. Fine Particulate Air Pollution and Daily Mortality. A Nationwide Analysis in 272 Chinese Cities. Am J Respir Crit Care Med 196:73–81; doi:10.1164/rccm.201609-1862OC.

Crouse DL, Philip S, Donkelaar A van, Martin RV, Jessiman B, Peters PA, et al. 2016. A New Method to Jointly Estimate the Mortality Risk of Long-Term Exposure to Fine Particulate Matter and its Components. Sci Rep 6:1–10; doi:10.1038/srep18916.

Egozcue JJ, Pawlowsky-Glahn V, Mateu-Figueras G, Barceló-Vidal C. 2003. Isometric Logratio Transformations for Compositional Data Analysis. Math Geol 35:279–300; doi:10.1023/A:1023818214614.

Florczyk AJ, Melchirorri M, Corbane C, Schiavina M, Maffenini M, Pesaresi M, et al. 2019. Description of the GHS Urban Centre Database 2015 : public release 2019 : version 1.0. Publ Off Eur Union; doi:10.2760/037310.

Franklin M, Koutrakis P, Schwartz J. 2008. The Role of Particle Composition on the Association Between PM2.5 and Mortality. Epidemiology 19: 680–689.

Franklin M, Zeka A, Schwartz J. 2007. Association between PM2.5 and all-cause and specific-cause mortality in 27 US communities. J Expo Sci Environ Epidemiol 17:279–287; doi:10.1038/sj.jes.7500530.

Gasparrini A. 2011. Distributed lag linear and non-linear models in R: the package dlnm. J Stat Softw 43: 1.

Gasparrini A, Armstrong B, Kenward MG. 2012. Multivariate meta-analysis for non-linear and other multi-parameter associations. Stat Med 31; doi:10.1002/sim.5471.

Giancristofaro RA, Gastaldi M, Martinello L, Meneguzzer C. 2020. Regression analysis with compositional data using orthogonal log-ratio coordinates. Commun Stat-Simul Comput; doi:10.1080/03610918.2019.1691224.

Hashizume M, Kim Y, Ng CFS, Chung Y, Madaniyazi L, Bell ML, et al. 2020. Health Effects of Asian Dust: A Systematic Review and Meta-Analysis. Environ Health Perspect 128:066001; doi:10.1289/EHP5312.

Higgins JPT, Thompson SG. 2002. Quantifying heterogeneity in a meta-analysis. Stat Med 21:1539–1558; doi:10.1002/sim.1186.

Hron K, Filzmoser P, Thompson K. 2012. Linear regression with compositional explanatory variables. J Appl Stat 39:1115–1128; doi:10.1080/02664763.2011.644268.

Huang W, Cao J, Tao Y, Dai L, Lu S-E, Hou B, et al. 2012. Seasonal Variation of Chemical Species Associated With Short-Term Mortality Effects of PM2.5 in Xi'an, a Central City in China. Am J Epidemiol 175:556–566; doi:10.1093/aje/kwr342.

Hvidtfeldt UA, Geels C, Sørensen M, Ketzel M, Khan J, Tjønneland A, et al. 2019. Long-term residential exposure to PM2.5 constituents and mortality in a Danish cohort. Environ Int 133:105268; doi:10.1016/j.envint.2019.105268.

Janssen NAH, Hoek G, Simic-Lawson M, Fischer P, van Bree L, ten Brink H, et al. 2011. Black Carbon as an Additional Indicator of the Adverse Health Effects of Airborne Particles Compared with PM10 and PM2.5. Environ Health Perspect 119:1691–1699; doi:10.1289/ehp.1003369.

Kelly FJ, Fussell JC. 2012. Size, source and chemical composition as determinants of toxicity attributable to ambient particulate matter. Atmos Environ 60:504–526; doi:10.1016/j.atmosenv.2012.06.039.

Kioumourtzoglou M-A, Austin E, Koutrakis P, Dominici F, Schwartz J, Zanobetti A. 2015. PM2.5 and survival among older adults: Effect modification by particulate composition. Epidemiology 26:321–327; doi:10.1097/EDE.0000000000000269.

Li Y, Henze DK, Jack D, Henderson BH, Kinney PL. 2016. Assessing public health burden associated with exposure to ambient black carbon in the United States. Sci Total Environ 539:515–525; doi:10.1016/j.scitotenv.2015.08.129.

Lin H, Tao J, Du Y, Liu T, Qian Z, Tian L, et al. 2016. Particle size and chemical constituents of ambient particulate pollution associated with cardiovascular mortality in Guangzhou, China. Environ Pollut 208:758–766; doi:10.1016/j.envpol.2015.10.056.

Liu C, Chen R, Sera F, Vicedo-Cabrera AM, Guo Y, Tong S, et al. 2019. Ambient Particulate Air Pollution and Daily Mortality in 652 Cities. N Engl J Med 381:705–715; doi:10.1056/NEJMoa1817364.

Liu S, Zhang K. 2015. Fine particulate matter components and mortality in Greater Houston: Did the risk reduce from 2000 to 2011? Sci Total Environ 538:162–168; doi:10.1016/j.scitotenv.2015.08.037.

Luben TJ, Nichols JL, Dutton SJ, Kirrane E, Owens EO, Datko-Williams L, et al. 2017. A systematic review of cardiovascular emergency department visits, hospital admissions and mortality associated with ambient black carbon. Environ Int 107:154–162; doi:10.1016/j.envint.2017.07.005.

Maraut S, Dernis H, Webb C, Spiezia V, Guellec D. 2008. The OECD REGPAT Database. Work Pap.

Martín-Fernández JA, Barceló-Vidal C, Pawlowsky-Glahn V. 2003. Dealing with Zeros and Missing Values in Compositional Data Sets Using Nonparametric Imputation. Math Geol 35:253–278; doi:10.1023/A:1023866030544.

Martín-Fernández JA, Hron K, Templ M, Filzmoser P, Palarea-Albaladejo J. 2012. Model-based replacement of rounded zeros in compositional data: Classical and robust approaches. Comput Stat Data Anal 56:2688–2704; doi:10.1016/j.csda.2012.02.012.

McDuffie EE, Smith SJ, O'Rourke P, Tibrewal K, Venkataraman C, Marais EA, et al. 2020. A global anthropogenic emission inventory of atmospheric pollutants from sector- and fuel-specific sources (1970–2017): An application of the Community Emissions Data System (CEDS). Earth Syst Sci Data Discuss 1–49; doi:https://doi.org/10.5194/essd-2020-103.

Meng J, Martin RV, Li C, van Donkelaar A, Tzompa-Sosa ZA, Yue X, et al. 2019. Source Contributions to Ambient Fine Particulate Matter for Canada. Environ Sci Technol 53:10269–10278; doi:10.1021/acs.est.9b02461.

Mert MC, Filzmoser P, Endel G, Wilbacher I. 2016. Compositional data analysis in epidemiology. Stat Methods Med Res 27:1878–1891; doi:10.1177/0962280216671536.

Mostofsky E, Schwartz J, Coull BA, Koutrakis P, Wellenius GA, Suh HH, et al. 2012. Modeling the Association Between Particle Constituents of Air Pollution and Health Outcomes. Am J Epidemiol 176:317–326; doi:10.1093/aje/kws018.

Muller I, Hron K, Fiserova E, Smahaj J, Cakirpaloglu P, Vancakova J. 2018. Interpretation of Compositional Regression with Application to Time Budget Analysis. Austrian J Stat 47:3–19; doi:10.17713/ajs.v47i2.652.

Ostro B, Lipsett M, Reynolds P, Goldberg D, Hertz A, Garcia C, et al. 2010. Long-Term Exposure to Constituents of Fine Particulate Air Pollution and Mortality: Results from the California Teachers Study. Environ Health Perspect 118:363–369; doi:10.1289/ehp.0901181.

Palarea-Albaladejo J, Martín-Fernández JA. 2015. zCompositions — R package for multivariate imputation of left-censored data under a compositional approach. Chemom Intell Lab Syst 143:85–96; doi:10.1016/j.chemolab.2015.02.019.

Park RJ, Jacob DJ, Field BD, Yantosca RM, Chin M. 2004. Natural and transboundary pollution influences on sulfate-nitrate-ammonium aerosols in the United States: Implications for policy. J Geophys Res Atmospheres 109; doi:10.1029/2003JD004473.

Pearson K. 1897. Mathematical contributions to the theory of evolution.—On a form of spurious correlation which may arise when indices are used in the measurement of organs. Proc R Soc Lond 60:489–498; doi:10.1098/rspl.1896.0076.

Peng RD, Bell ML, Geyh AS, McDermott A, Zeger SL, Jonathan M. S, et al. 2009. Emergency Admissions for Cardiovascular and Respiratory Diseases and the Chemical Composition of Fine Particle Air Pollution. Environ Health Perspect 117:957–963; doi:10.1289/ehp.0800185.

Philip S, Martin RV, Snider G, Weagle CL, Donkelaar A van, Brauer M, et al. 2017. Anthropogenic fugitive, combustion and industrial dust is a significant, underrepresented fine particulate matter source in global atmospheric models. Environ Res Lett 12:044018; doi:10.1088/1748-9326/aa65a4.

Pinder RW, Adams PJ, Pandis SN. 2007. Ammonia Emission Controls as a Cost-Effective Strategy for Reducing Atmospheric Particulate Matter in the Eastern United States. Environ Sci Technol 41:380–386; doi:10.1021/es060379a.

R Core Team. 2020. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing:Vienna, Austria.

Rückerl R, Schneider A, Breitner S, Cyrys J, Peters A. 2011. Health effects of particulate air pollution: A review of epidemiological evidence. Inhal Toxicol 23:555–592; doi:10.3109/08958378.2011.593587.

Sera F, Armstrong B, Blangiardo M, Gasparrini A. 2019a. An extended mixed-effects framework for meta-analysis. Stat Med 38:5429–5444; doi:10.1002/sim.8362.

Sera F, Armstrong B, Tobias A, Vicedo-Cabrera AM, Åström C, Bell ML, et al. 2019b. How urban characteristics affect vulnerability to heat and cold: a multi-country analysis. Int J Epidemiol; doi:10.1093/ije/dyz008.

Son J-Y, Lee J-T, Kim K-H, Jung K, Bell ML. 2012. Characterization of Fine Particulate Matter and Associations between Particulate Chemical Constituents and Mortality in Seoul, Korea. Environ Health Perspect 120:872–878; doi:10.1289/ehp.1104316.

Stafoggia M, Zauli-Sajani S, Pey J, Samoli E, Alessandrini E, Basagaña X, et al. 2016. Desert Dust Outbreaks in Southern Europe: Contribution to Daily PM10 Concentrations and Short-Term Associations with Mortality and Hospital Admissions. Environ Health Perspect 124:413–419; doi:10.1289/ehp.1409164.

Stanaway JD, Afshin A, Gakidou E, Lim SS, Abate D, Abate KH, et al. 2018. Global, regional, and national comparative risk assessment of 84 behavioural, environmental and occupational, and metabolic risks or clusters of risks for 195 countries and territories, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017. The Lancet 392:1923–1994; doi:10.1016/S0140-6736(18)32225-6.

van den Boogaart KG, Tolosana-Delgado R. 2008. “compositions”: A unified R package to analyze compositional data. Comput Geosci 34:320–338; doi:10.1016/j.cageo.2006.11.017.

van Donkelaar A, Martin RV, Li C, Burnett RT. 2019. Regional Estimates of Chemical Composition of Fine Particulate Matter Using a Combined Geoscience-Statistical Method with Information from Satellites, Models, and Monitors. Environ Sci Technol 53:2595–2611; doi:10.1021/acs.est.8b06392.

Wang M, Beelen R, Stafoggia M, Raaschou-Nielsen O, Andersen ZJ, Hoffmann B, et al. 2014. Long-term exposure to elemental constituents of particulate matter and cardiovascular mortality in 19 European cohorts: Results from the ESCAPE and TRANSPHORM projects. Environ Int 66:97–106; doi:10.1016/j.envint.2014.01.026.

Wu Y, Gu B, Erisman JW, Reis S, Fang Y, Lu X, et al. 2016. PM2.5 pollution is substantially affected by ammonia emissions in China. Environ Pollut 218:86–94; doi:10.1016/j.envpol.2016.08.027.

Yang Y, Ruan Z, Wang X, Yang Y, Mason TG, Lin H, et al. 2019. Short-term and long-term exposures to fine particulate matter constituents and health: A systematic review and meta-analysis. Environ Pollut 247:874–882; doi:10.1016/j.envpol.2018.12.060.

Tables

Table 1: Description of first-stage data aggregated per country

Country	Cities	Data period*	Total mortality	Mean $PM_{2.5}$ (10 – 90 percentiles) in µg/m^3
Australia	3	2000-2009	388 122	7.0 (3.2 – 11.9)
Canada	18	1999-2015	1 767 732	8.1 (2.6 – 15.2)
Chile	3	2008-2014	265 084	34.2 (8.7 – 64.7)
China	3	2013-2015	248 716	61.2 (19.9 – 120.4)
Estonia	1	2008-2015	8 226	9.6 (2.1 – 19.4)
Finland	1	1999-2014	117 610	16.8 (4.8 – 34.4)
Germany	11	2004-2015	1 051 813	15.4 (5.6 - 29.0)
Greece	1	2007-2010	118 034	21.9 (11.5 – 34.0)
Japan	35	2011-2015	1 292 348	14.3 (5.5 – 25.5)
Mexico	3	2003-2012	1 148 573	27.0 (14.0 – 41.3)
Portugal	1	2004-2018	315 615	12.5 (4.9 – 23.2)
South Africa	1	2004-2013	322 999	37.4 (16.6 - 64.0)
Spain	12	2009-2013	229 992	11.6 (4.9 - 20.1)
Sweden	1	2001-2010	90 670	8.2 (3.6 - 14.4)
Switzerland	4	1999-2013	128 779	19.3 (6.7 - 35.8)
Taiwan	2	2007-2014	369 048	30.5 (13.8 - 51.5)
UK	24	1999-2016	1 556 506	12.3 (4.8 - 23.3)
USA	78	1999-2006	5 251 542	13.1 (5.1 - 23.5)

* For the first stage only. It may slightly vary within countries because of missing values.

Table 2: Measures of residual heterogeneity for nested fixed-effects specifications.

	Cochran Q	I^2 (%)	Wald statistic*	p-value*
Full model	318.3	39.4	16.7	0.0103
Only indicator PC	473.5	58.0	0.4	0.8111
No fixed effect	488.2	58.8	-	-

* Wald statistic and associated p-value test nested hypotheses compared to the model on the line below.

Figures

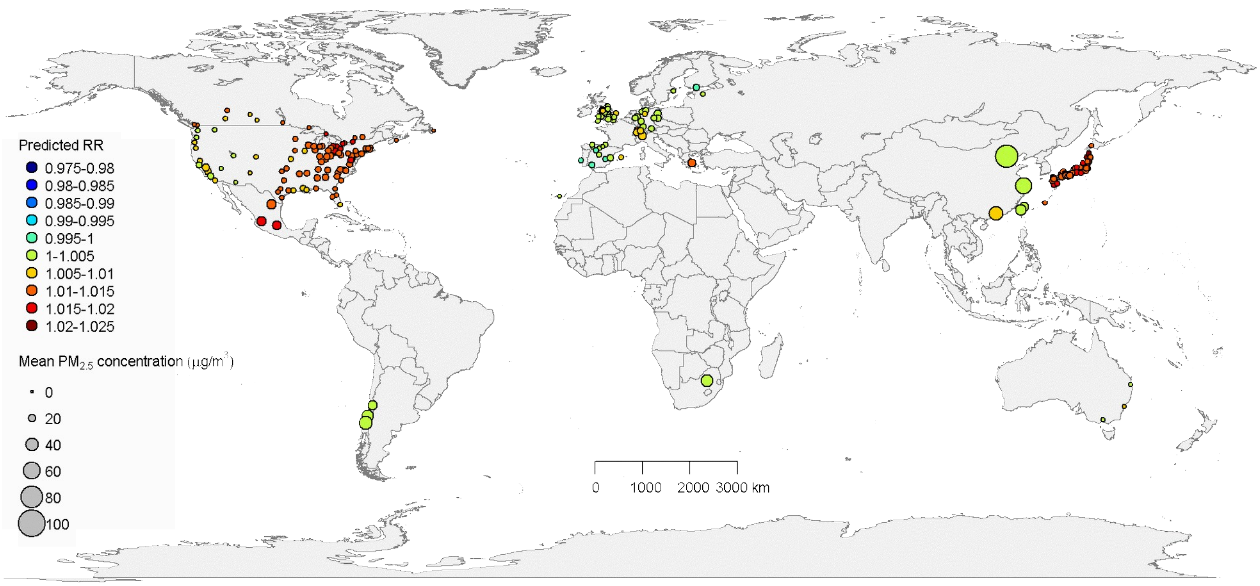

Figure 1: Locations used in the study with their mean PM2.5 concentration and best linear unbiased predictions (BLUPs) of relative risks (RRs) for mortality.

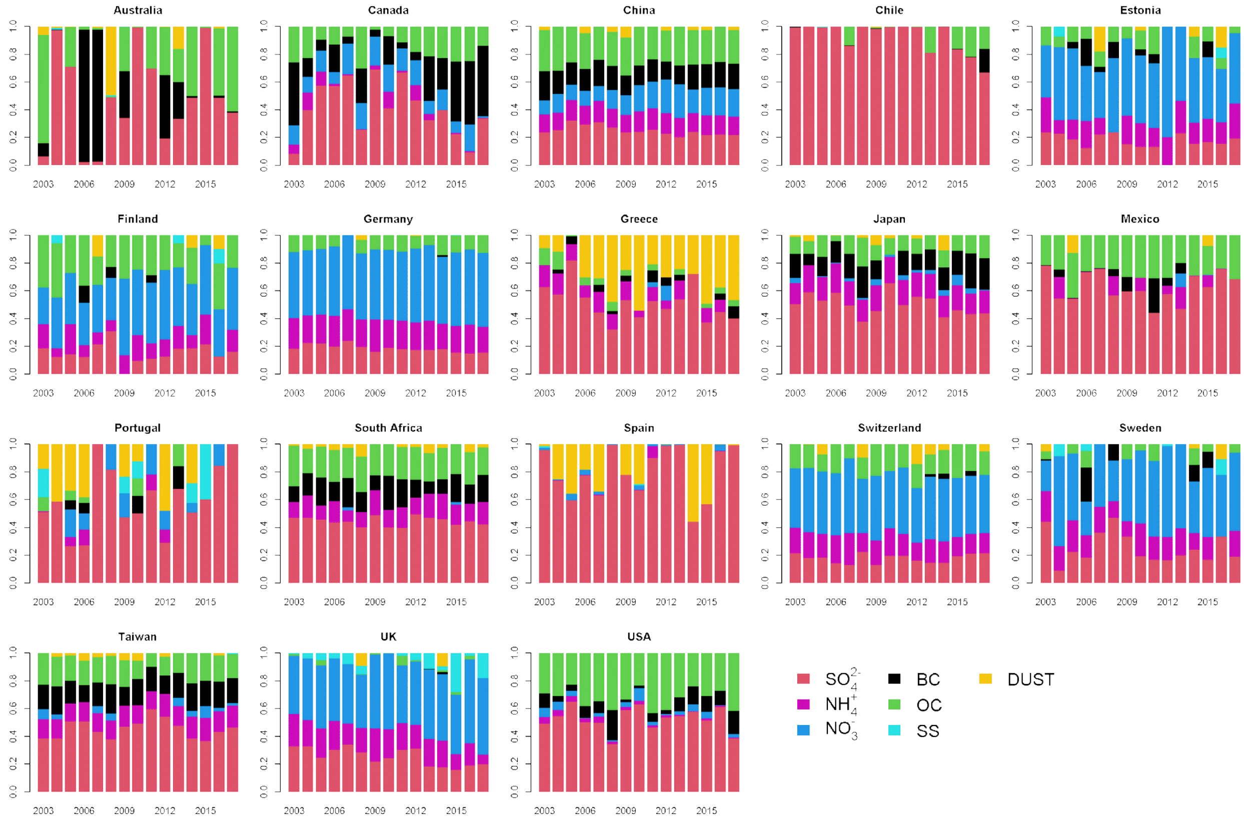

Figure 2: Geometrical mean of the $PM^{2.5}$ composition in each country.

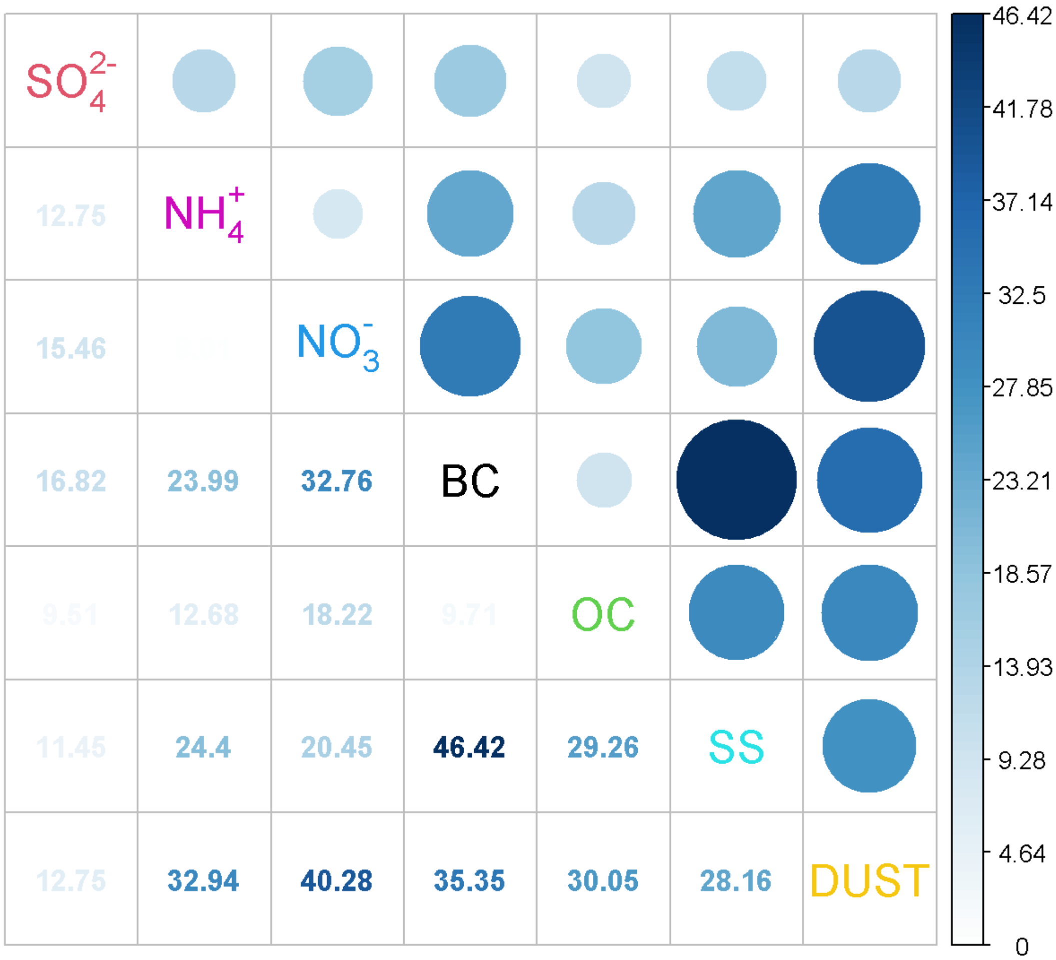

Figure 3: Variation matrix of the $PM_{2.5}$ composition. The upper side colour and circle size represent the values displayed on the lower side of the diagonal. A large variation value indicates that components tend to vary against each other.

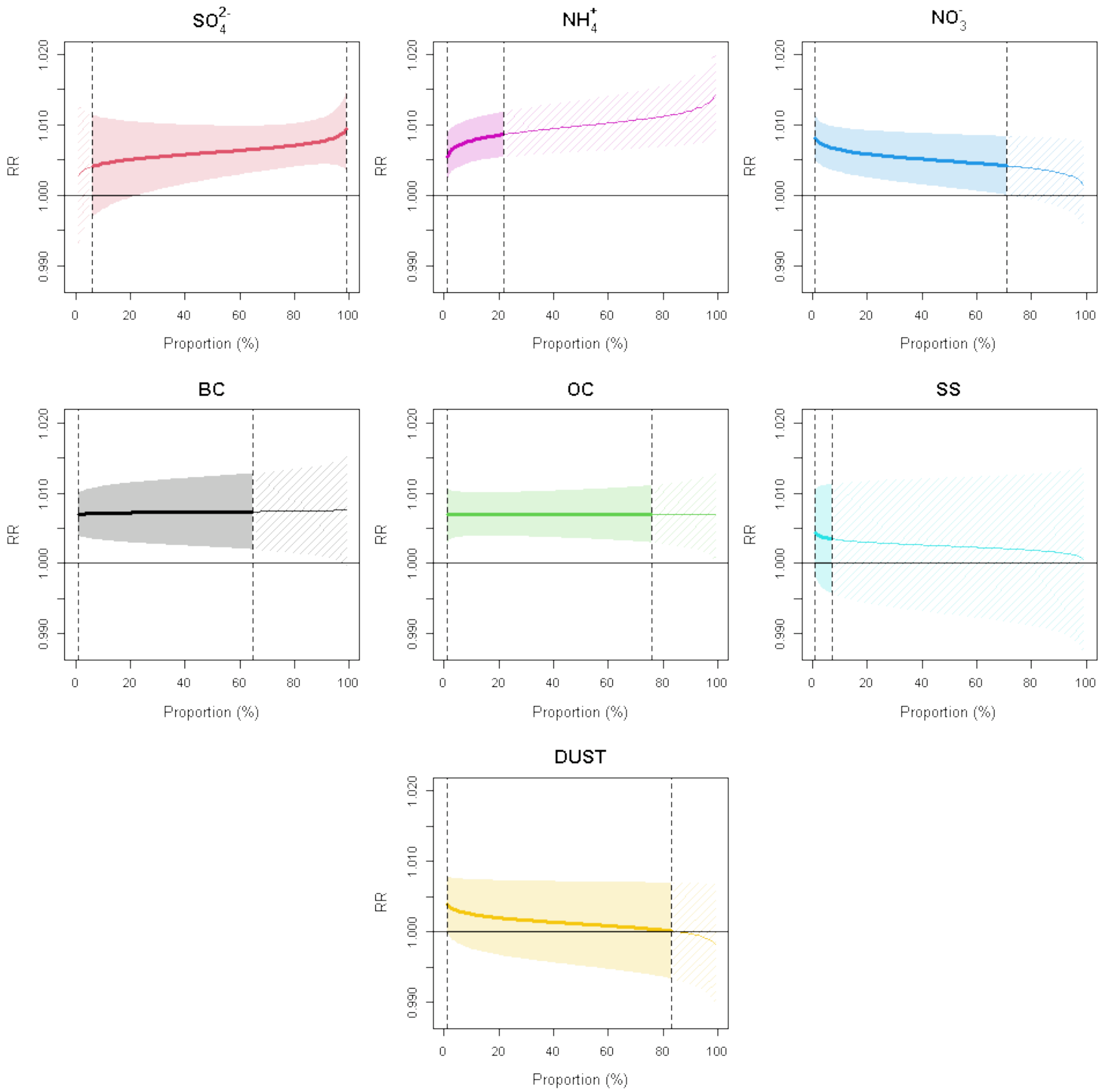

Figure 4: Predicted relative risks (RRs) for different values of each component while keeping the other constituents constant. The predicted RR is associated with an increase of 10µg/m^3 in PM$_{2.5}$. Thick lines indicate the range of observed values for each component, while thin dashed lines indicate extrapolations. Coloured bands represent 95% confidence regions.