Investigating the health effect of multi-pollutant exposure using a time series approach

Marta Blangiardo

Imperial College London
MRC-PHE Centre for Environment and Heath
m.blangiardo@imperial.ac.uk

TIES Conference
Bergamo University
24-26 July 2017

MRC-PHE
Centre for Environment & Health

Air pollution

- The air we breath includes a complex mixture of thousands of pollutants; solid or liquid particles as well as gases.

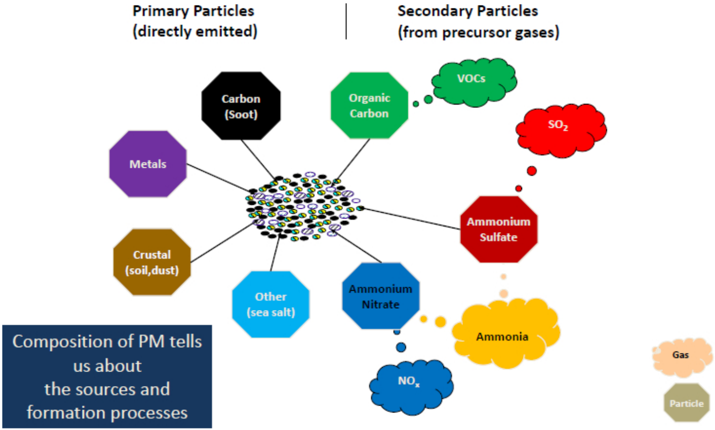


Image credit: *Gary W. Fuller*

Health effects of polluted air: some estimates

- IARC (2013): Outdoor air pollution is carcinogenic to humans;
- WHO/Europe (2006): Outdoor PM causes a reduction in life expectancy of average population by approximately a year in Europe;
- WHO (2016): Outdoor $PM_{2.5}$ causes more than 3 million deaths per year worldwide; 92% of the world's population lives in places where air quality levels exceed WHO limits.

BBC Sign in News Sport Weather iPlayer

NEWS

Home UK World Business Politics Tech Science Health Education

Health

UK air pollution 'linked to 40,000 early deaths a year'

By Michelle Roberts
Health editor, BBC News online

23 February 2016 | Health

Outdoor air pollution is contributing to about 40,000 early deaths a year in the UK, say the Royal Colleges of Physicians and of Paediatrics and Child Health.

theguardian

UK election world sport football opinion culture business lifestyle fashion environment tech travel ≡ browse all sections

home › environment › pollution climate change wildlife energy

Pollution

Pollution responsible for quarter of deaths of young children, says WHO

Toxic air, unsafe water and and lack of sanitation cause the deaths of 1.7 million under-fives every year

This article is 3 months old

7,339 191

Damian Carrington
@dpcarrington

Tuesday 7 March 2017 11.43 GMT

Children wear masks to protect against pollution at a school in Beijing, China. Photograph: Greg Baker/AFP/Getty Images

Pollution is responsible for one in four deaths among all children under five, according to new World Health Organisation reports, with toxic air, unsafe water, and and lack of sanitation the leading causes.

Short-term vs Long-term effects

- Air pollution effects on health can be **short** or **long** term;
- Short term considers the day-to-day variation in air pollution and evaluate the association with the day-to-day variation in health:
  - typically air pollution data from monitoring stations (one or more);
  - important to account for seasonality and meteorological variables;
  - statistical modelling usually in a time-series framework.
- Long term considers (yearly) averages of air pollution concentration and counts of outcomes (at individual level or) over small areas:
  - typically air pollution data from exposure models (deterministic or more recently statistical);
  - important to account for area level confounders (e.g. population and area characteristics);
  - statistical approach is usually ecological regression.

From a single to a multi-pollutant approach

- The quantification of the impact of air pollution on population health has been historically undertaken through a **single pollutant approach**;
- This is mainly due to:
  - measurement and source complexities which have limited the development of statistically robust multi-pollutant models;
  - regulatory strategies of air quality management which have addressed a single pollutant at a time.

However, the air we breathe is a mixture and:

- It is unlikely that all parts of the air pollution mix are equally harmful;

Therefore, we need new/revised statistical methods and approaches for a multi-pollutant approach (e.g. Coull and Park (2015) and Molitor et al. (2016)).

In this talk

1. Present a multi-pollutant hierarchical model which jointly estimate pollutant concentration and their link with health.

2. Application on six pollutants and daily mortality in Greater London in 2011-2012.

Multi-pollutant approach so far

- Air quality indexes (Daily Air Quality Index in the UK)
  → typically used by governments;
  → easy to build and to communicate to the public.

- Bayesian Kernel Machine Regression (Bobb et al. 2015)
  → the pollutants are included in the model through a smooth function represented using a kernel;
  → authors found that Gaussian kernel outperformed linear and ridge regression kernels.

- Dirichlet process mixture model (profile regression, (Molitor et al. 2010))
  → days are clustered based on their concentration profiles;
  → concentration and health outcome are modelled jointly.

Another way forward

We propose the use of a hierarchical Bayesian time-series approach which is formed by two linked components:

- A **pollutant component** which estimates the true 'latent' concentration values;
- A **health component** which links the estimated concentration to the health outcome;
- The two components are jointly modelled;
- The modelling framework allows to estimate a health effect for each pollutant.

Data description

- Daily measurements of five regulated pollutants and the number of particles present in any given volume of air (PCN) available from a monitoring site in North Kensington for 2011-2012.
- Daily count of mortality for cardio-vascular disease (ICD-10, Chapter I) available for the same period.

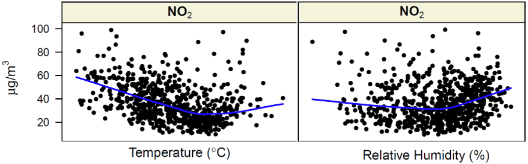
	Number of Days	Percentiles 10th	25th	50th	75th	90th	IQR
Mortality	731	28	32	37	42	47	10
Meteorological data:							
Temperature ($^\circ C$)	731	5.1	8.0	11.7	15.5	18.1	7.4
Relative Humidity (%)	731	61.6	69.6	78.0	84.2	88.5	14.5
Pollutants:							
CO (mg/m^3)	715	0.1	0.2	0.2	0.3	0.4	0.1
NO_2 ($\mu g/m^3$)	706	18.2	23.2	33.3	46.9	57.9	23.6
O_3 ($\mu g/m^3$)	695	11.4	24.3	39.1	51.1	64.9	26.8
SO_2 ($\mu g/m^3$)	717	0.0	0.4	1.8	2.6	3.6	2.2
$PM_{2.5}$ ($\mu g/m^3$)	730	5.0	6.0	9.0	14.0	25.0	8.0
PCN (p/mm^3)	636	7.8	9.7	12.1	14.9	17.9	5.2

H2M: POLLUTANT COMPONENT

- We specify x_{pt} as the measured (standardised) concentration level of pollutant p ($p = 1, ..., P = 6$) on day t ($t = 1, ..., T = 731$) from the monitoring site:

$$
\begin{aligned}
x_{pt} &\sim N(\mu_{pt}, \sigma_p^2) && \text{Measurement error model} \\
\mu_{pt} &= \gamma_{0p} + \sum_j \gamma_{jp} \mathrm{z}_{jt} + \theta_{pt} && \textbf{True} \text{ (latent) concentration model}
\end{aligned}
$$

- $\mathbf{z}_t$ - meteorological variables:



$\Rightarrow$ Evidence of non-linear relationship with the pollutant levels - inclusion of linear and quadratic terms.

H2M: pollutant component - θ

- $\{\theta_{1t}, \dots, \theta_{Pt}\}$ accounts for the residual temporal effects and for the correlation among pollutants

$$(\theta_{1t}, ..., \theta_{Pt})^T \sim \text{MVN}\left((\theta_{1,t-\ell}, ..., \theta_{P,t-\ell})^T, \Sigma_P\right)$$

- $t - \ell$ provides the temporal lag of ℓ days for the t-th day;
- The diagonal of the covariance matrix of the errors Σ_P allows each pollutant to have a different amount of temporal dependence;
- The off-diagonals represent the temporal dependence between the pollutants.

H2M: HEALTH COMPONENT

- The second model component links the **true** concentration μ_{pt} to the health outcome

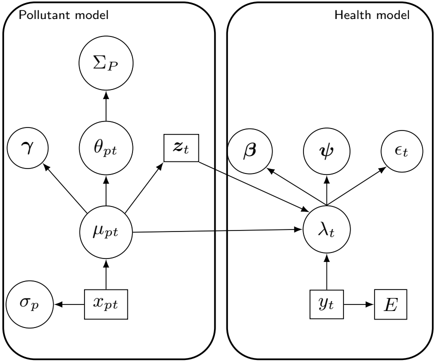
$$
\begin{aligned}
y_t &\sim \text{Poisson}(\lambda_t E) \\
\log(\lambda_t) &= \beta_0 + \sum_p \beta_p \mu_{p(t-1)} + \sum_i s(v_{ti}, \psi_i) + \delta_{I_t} + \epsilon_t
\end{aligned}
$$

- λ_t represents the relative risk of CVD death on day t compared to the average;
- we consider lag $\ell = 1$ (same as Atkinson et al., 2016);
- $\boldsymbol{\beta}$ are the pollutant effects on CVD mortality;
- ϵ_t is an overdispersion parameter;
- $s(v_{ti}, \psi_i)$ is modelled through a low-rank thin plate spline:

$$
s(v_{ti}, \psi_i) = \alpha_i v_{ti} + \sum_{k=1}^{K_i} b_{ki} |v_{ti} - \kappa_{ki}|^3
$$

H2M: Graphical representation

- Pollutant and health components are jointly modelled
  $\Rightarrow$ Uncertainty on μ_{pt} is fed forward
  $\Rightarrow$ Information on the outcome y_t is fed back



- $\Sigma_P \sim \text{IW}(D, d)$ with $d = P$
- Regression coefficients are $\text{N}(0, 10^5)$
- Measurement error sd $\sigma_p \sim \text{U}(0, 100)$
- Coefficients for the basis functions: $b_{ki} \sim \text{N}(0, \sigma^2_{b_i})$; $1/\sigma^2_{b_i} \sim \text{Ga}(1, 0.01)$

Simulation study: setting

We consider (i) 6 pollutants, (ii) 2000 days, (iii) degree of correlation among pollutants varying between -0.6 and 0.8.
We generate:

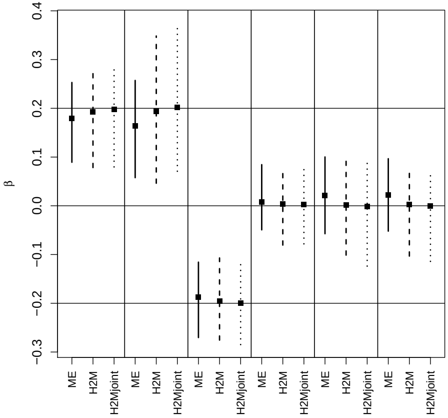
1. true pollutant concentration through $\mu_{pt} \sim \mathrm{N}(\mu_{p(t-1)}, \Sigma_P)$
2. measured concentration through $x_{pt} \sim \mathrm{N}(\mu_{pt}, 0.1)$
3. number of daily health events through a Poisson
4. β_p equal to -0.2, 0 or 0.2.
5. Joint vs two-stage model

We estimate our H2M framework and compare it against a simpler model which include the pollutant concentration with measurement error (ME):

$$\begin{aligned} y_t &\sim \text{Poisson}(\lambda_t E) \\ \log(\lambda_t) &= \beta_0 + \sum_p \beta_p x_{p(t-1)} + \epsilon_t \end{aligned}$$

Simulation study: results

- Bias and RMSE are substantially reduced for H2Ms;
- Coverage improves in H2Ms;
- Uncertainty is larger for H2Ms.



	Bias			RMSE			95% CI coverage		
	ME	H2Mj	H2M	ME	H2Mj	H2M	ME	H2Mj	H2M
β_1	-0.021	-0.002	-0.007	0.003	0.002	0.002	65	93	92
β_2	-0.036	0.002	-0.006	0.004	0.004	0.005	53	97	97
β_3	0.013	0.000	0.004	0.003	0.002	0.004	71	97	92
β_4	0.008	0.002	0.003	0.001	0.001	0.002	77	99	98
β_5	0.021	-0.001	0.002	0.002	0.002	0.002	61	94	95
β_6	0.022	-0.001	0.002	0.002	0.002	0.002	65	99	97

Application Results: single vs multi-pollutant model

- Five pollutants (CO, NO_2, O_3, SO_2, $PM_{2.5}$) and particle number concentration (PCN);
- NO_2 and O_3 are the only pollutants to show evidence of an effect on health;
- In comparison in the single pollutant approach the effects are pushed towards 0.

Pollutant		**IQR**	Multi-pollutant model **% Increase (95% CI)**		Single pollutant model **% Increase (95% CI)**	
CO	(mg/m^3)	0.10	-1.27	(-5.07, 2.54)	-1.74	(-4.01, 0.53)
NO_2	$(\mu g/m^3)$	23.65	**2.45**	**(1.18, 5.81)**	-0.53	(-3.13, 2.11)
O_3	$(\mu g/m^3)$	26.85	**3.84**	**(0.48, 7.18)**	2.71	(-0.09, 5.57)
SO_2	$(\mu g/m^3)$	2.20	-2.13	(-7.04, 2.82)	-1.39	(-5.17, 2.48)
PCNT	(p/mm^3)	5.18	-3.89	(-7.77, 0.22)	-0.52	(-3.74, 2.89)
$PM_{2.5}$	$(\mu g/m^3)$	8.00	-1.37	(-3.77, 0.95)	-0.82	(-2.09, 0.46)

Discussion points

- Modelling framework which allows to include more than one pollutant, accounting for their correlation;
  $\Rightarrow$ Can disclose combined effect which are not visible in a single-pollutant model.

- Uncertainty from the pollutant concentration estimates included in estimating the health effects; feedback from outcome is allowed;
  $\Rightarrow$ From the simulation study it seems this helps reduce the bias in the estimates.

- Next: Move to a spatio-temporal model.

Acknowledgements

- Lauren Kanapka
- Monica Pirani
- Gary Fuller

MRC-PHE
Centre for Environment & Health

Public Health England
MRC Medical Research Council
Imperial College London
King's College London

References

Bobb, J., Valeri, L., Henn, B., Christiani, D., Wright, R., Mazumdar, M., Godleski, J., and Coull, B. (2015), "Bayesian kernel machine regression for estimating the health effects of multi-pollutant mixtures," *Biostatistics*, 16, 493–508.

Coull, B. A. and Park, E. S. (2015), *Development of Statistical Methods for Multipollutant Research*, Boston, MA: Health Effects Institute, Research Report 183 - I & II.

IARC (2013), *IARC monographs on the evaluation of carcinogenic risks to humans. Volume 109. Outdoor air pollution*, Lyon, France: International Agency for Research on Cancer.

Molitor, J., Coker, E., Jerrett, M., Ritz, B., and Li, A. (2016), *Modeling of Multipollutant Profiles and Spatially Varying Health Effects with Applications to Indicators of Adverse Birth Outcomes*, Boston, MA: Health Effects Institute, Research Report 183 - III.

Molitor, J., Papathomas, M., Jerrett, M., and Richardson, S. (2010), "Bayesian profile regression with an application to the National Survey of Children's Health," *Biostatistics*, 11, 484–498.

WHO (2016), *Ambient air pollution: A global assessment of exposure and burden of disease*, Geneva, Switzerland.

WHO/Europe (2006), *Health Risks of Particulate Matter from Long-range Transboundary Air Pollution*, Copenhagen, Denmark: WHO Regional Office for Europe.