Identifying the Most Harmful Sources of Ambient Air Pollution to Better Protect Public Health

Jenna R. Krall

Postdoctoral Fellow
Department of Biostatistics and Bioinformatics
Emory University

Air pollution and health

Particulate matter (PM)

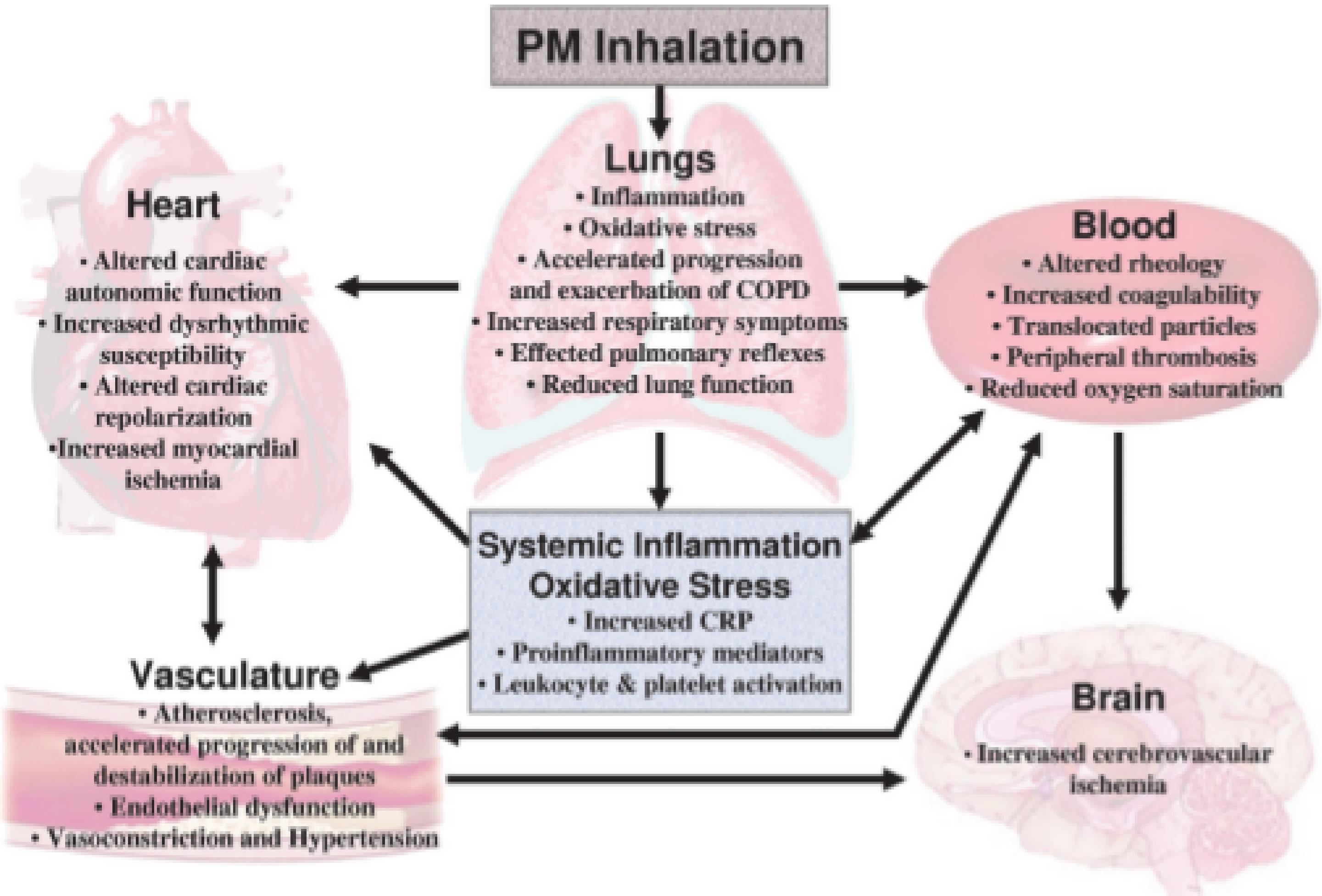


Image: Pope and Dockery (2006)

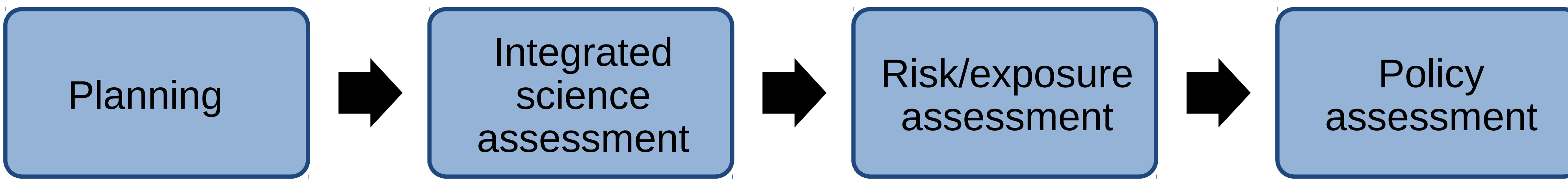
National Ambient Air Quality Standards (NAAQS)

Pollutant [links to historical tables of NAAQS reviews]		Primary/ Secondary	Averaging Time	Level	Form
Carbon Monoxide (CO)		primary	8 hours	9 ppm	Not to be exceeded more than once per year
			1 hour	35 ppm	
Lead (Pb)		primary and secondary	Rolling 3 month average	0.15 μg/m^3 (1)	Not to be exceeded
Nitrogen Dioxide (NO_2)		primary	1 hour	100 ppb	98th percentile of 1-hour daily maximum concentrations, averaged over 3 years
		primary and secondary	1 year	53 ppb (2)	Annual Mean
Ozone (O_3)		primary and secondary	8 hours	0.070 ppm (3)	Annual fourth-highest daily maximum 8-hour concentration, averaged over 3 years
Particle Pollution (PM)	$PM_{2.5}$	primary	1 year	12.0 μg/m^3	annual mean, averaged over 3 years
		secondary	1 year	15.0 μg/m^3	annual mean, averaged over 3 years
		primary and secondary	24 hours	35 μg/m^3	98th percentile, averaged over 3 years
	PM_{10}	primary and secondary	24 hours	150 μg/m^3	Not to be exceeded more than once per year on average over 3 years
Sulfur Dioxide (SO_2)		primary	1 hour	75 ppb (4)	99th percentile of 1-hour daily maximum concentrations, averaged over 3 years
		secondary	3 hours	0.5 ppm	Not to be exceeded more than once per year

Image: epa.gov

NAAQS review

Separately for each criteria pollutant:



- Individual pollutants vs. mixtures
- PM is a heterogeneous mixture
- Mixtures emitted from ***sources***
  - Interpretable
  - Better targets of intervention

Sources: where does air pollution come from?

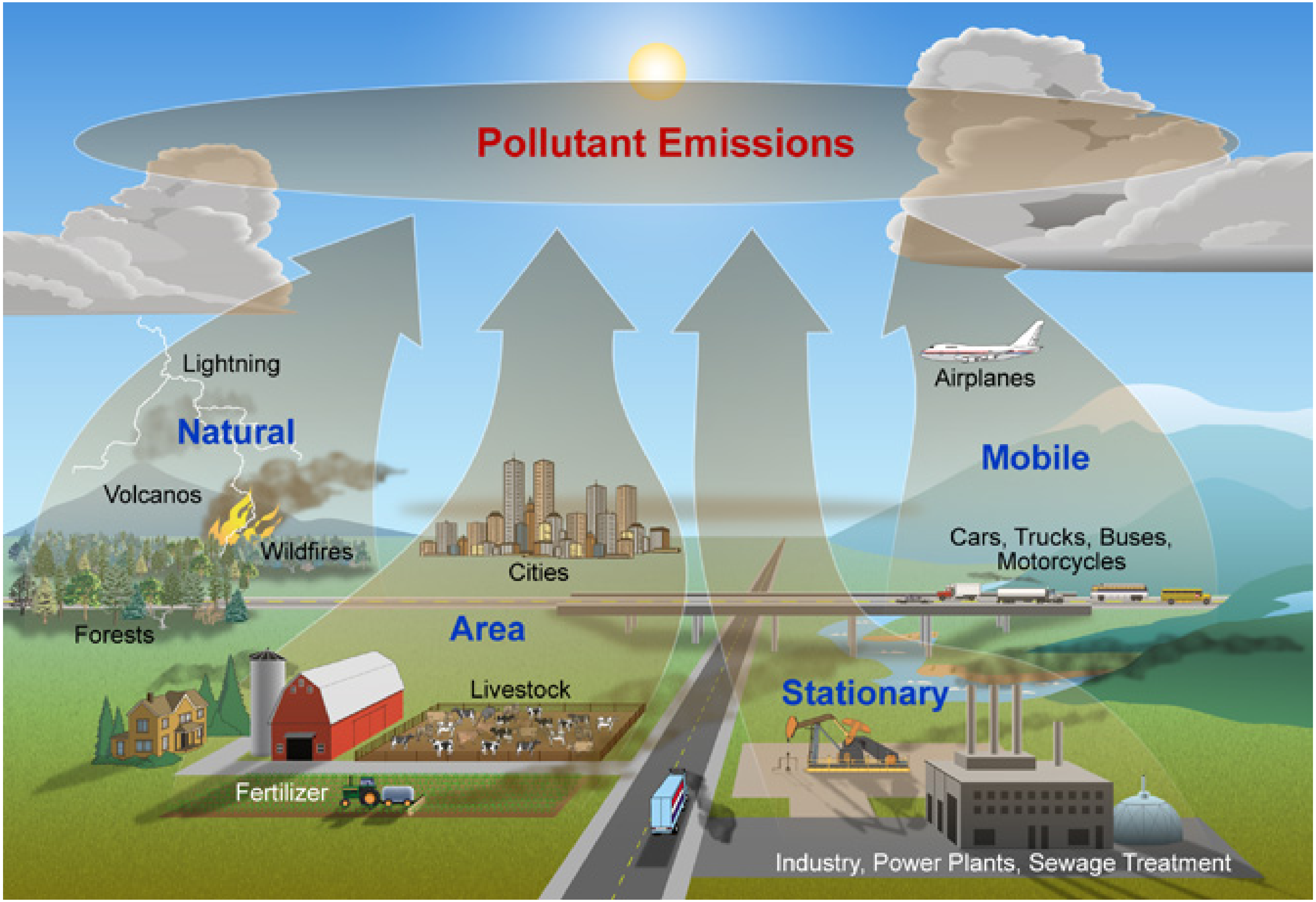
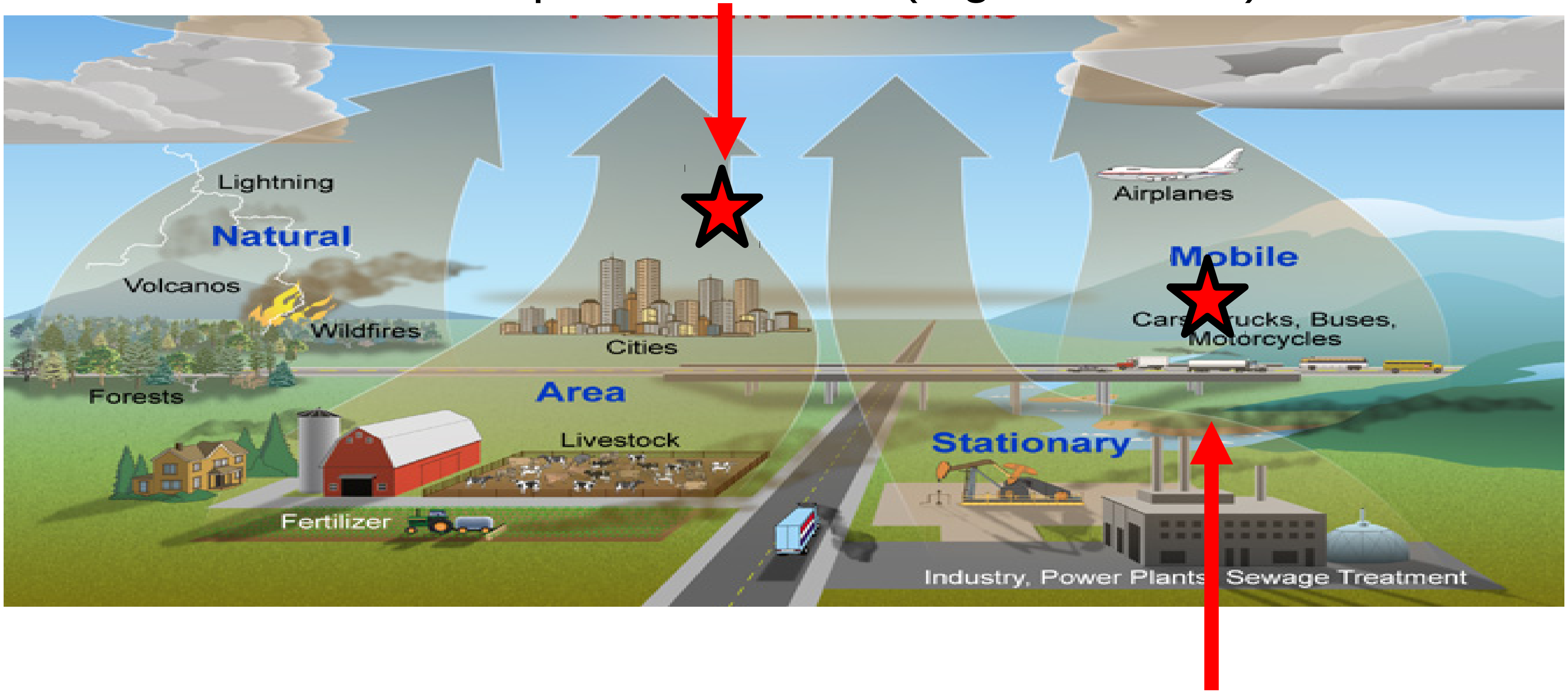


Image: nps.gov

Air pollution

Exposure-oriented air pollution monitor

- Ambient monitor
- Relevant exposure monitor (e.g. in-vehicle)



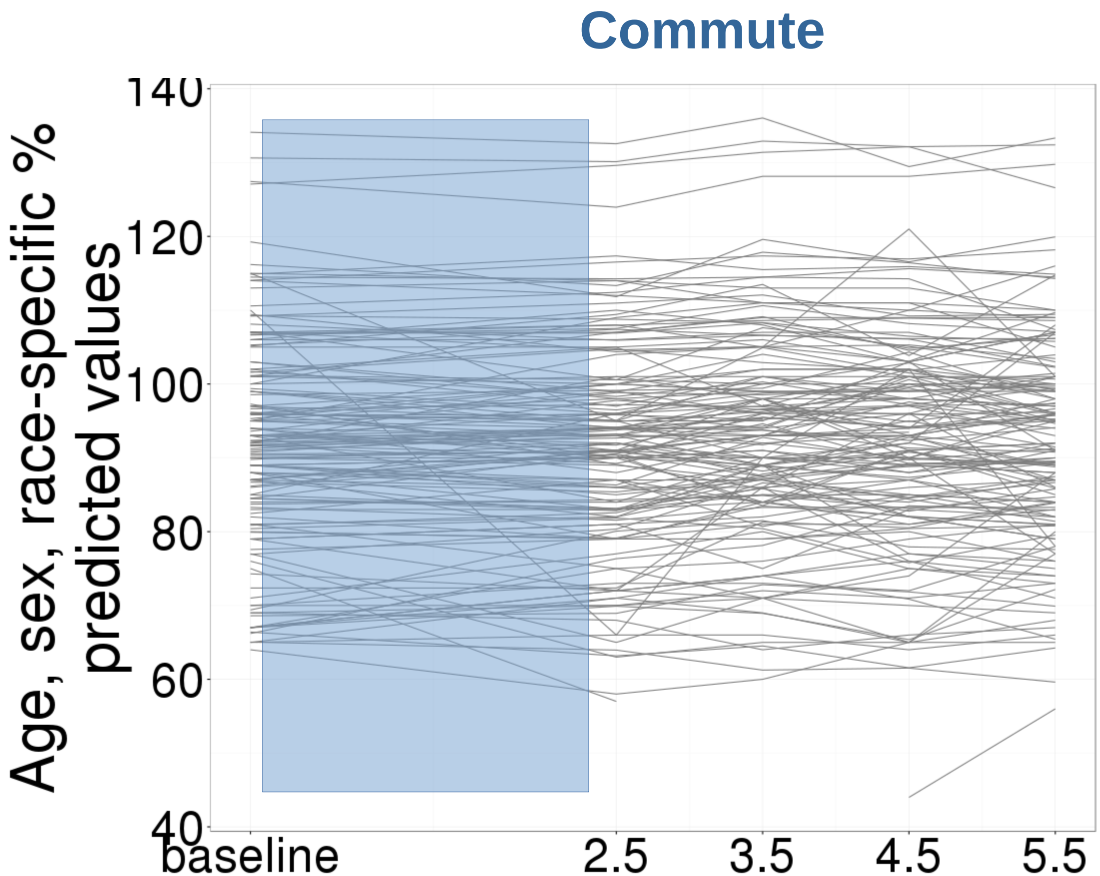
Source-oriented air pollution monitor

Image: nps.gov

Atlanta Commuters Exposure (ACE) Study

- 100 individuals performed 161 commutes around Atlanta, GA
- Healthy and asthmatic participants
- In-vehicle pollution measurements
- Pre and post-exposure lung function measurements

- Forced expiratory volume (FEV1)



Atlanta Commuters Exposure (ACE) Study

Subclinical health outcomes

- Forced expiratory volume in the first second (FEV1)
- Forced vital capacity (FVC)
- Exhaled nitric oxide (eNO)

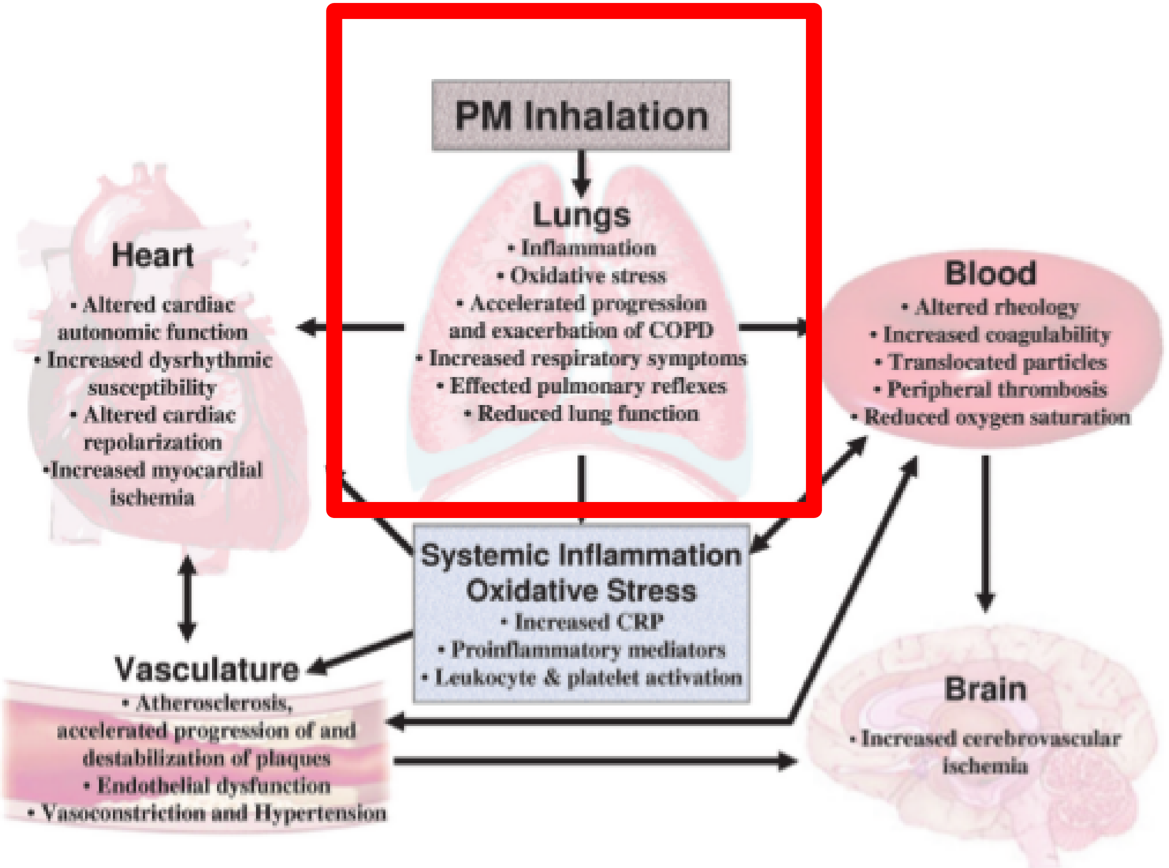


Image: Pope and Dockery (2006)

Atlanta Commuters Exposure (ACE) Study

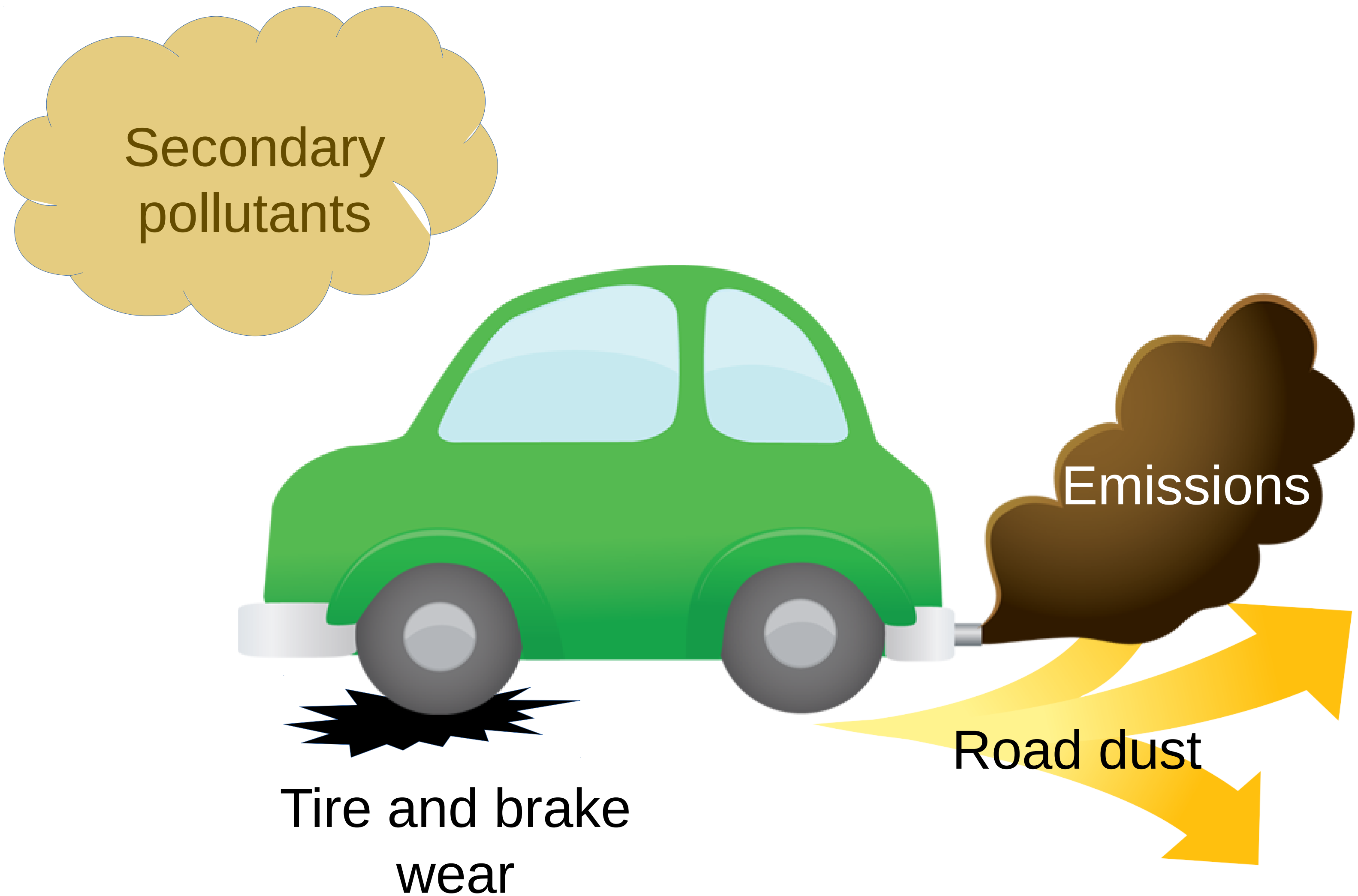


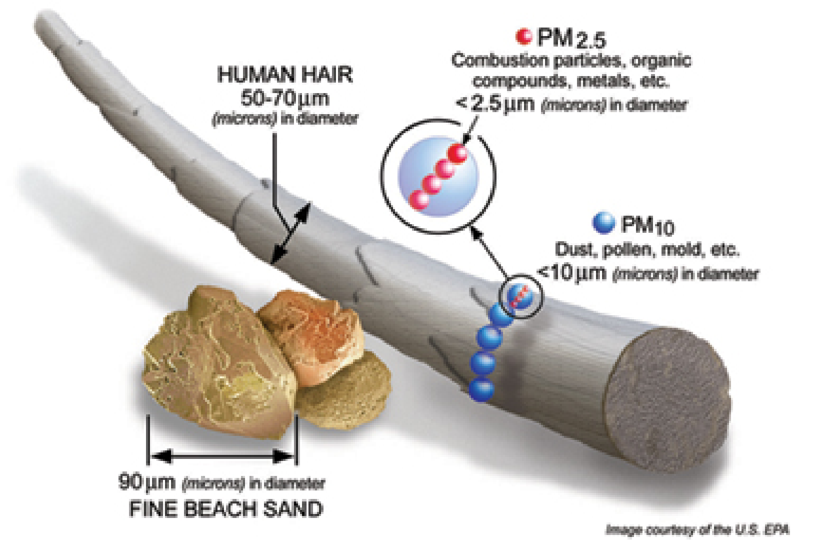
Image: LibreOffice

Atlanta Commuters Exposure (ACE) Study

PM pollutants

- Particulate matter (PM) less than 2.5 μm (PM2.5)
- Particle number concentration (PNC)
- Particle-bound polycyclic aromatic hydrocarbons (pbPAH)
- Chemical elements, including metals (Zn, Cd, Ni)
- Carbon-containing constituents
  - black carbon (BC)
  - Water-soluble organic carbon (OC)

Noise



Pollution data

Concentrations in μg/m³

Obs.	$PM_{2.5}$	Organic carbon	Elemental carbon	Nitrate	Sulfate	...
1	12.6	2.3	0.8	2.3	2.1	
2	15.7	3.8	1.2	2.7	2.6	
3	13.9	3.1	1.2	2.2	3.3	
4	14.5	4	1.7	2.7	3.3	
5	16.6	5.4	1.9	2.1	2.8	
6	8.6	1.7	0.5	0.5	2.9	
7	8.4	1.4	0.6	0.9	3	
8	10	2	1.1	1.1	2.2	
...						

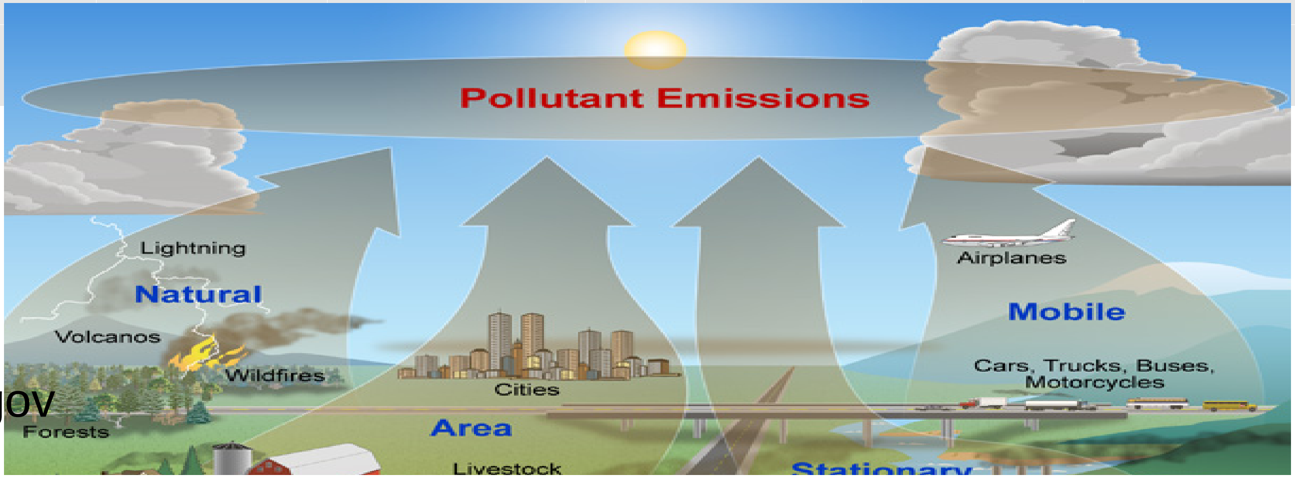
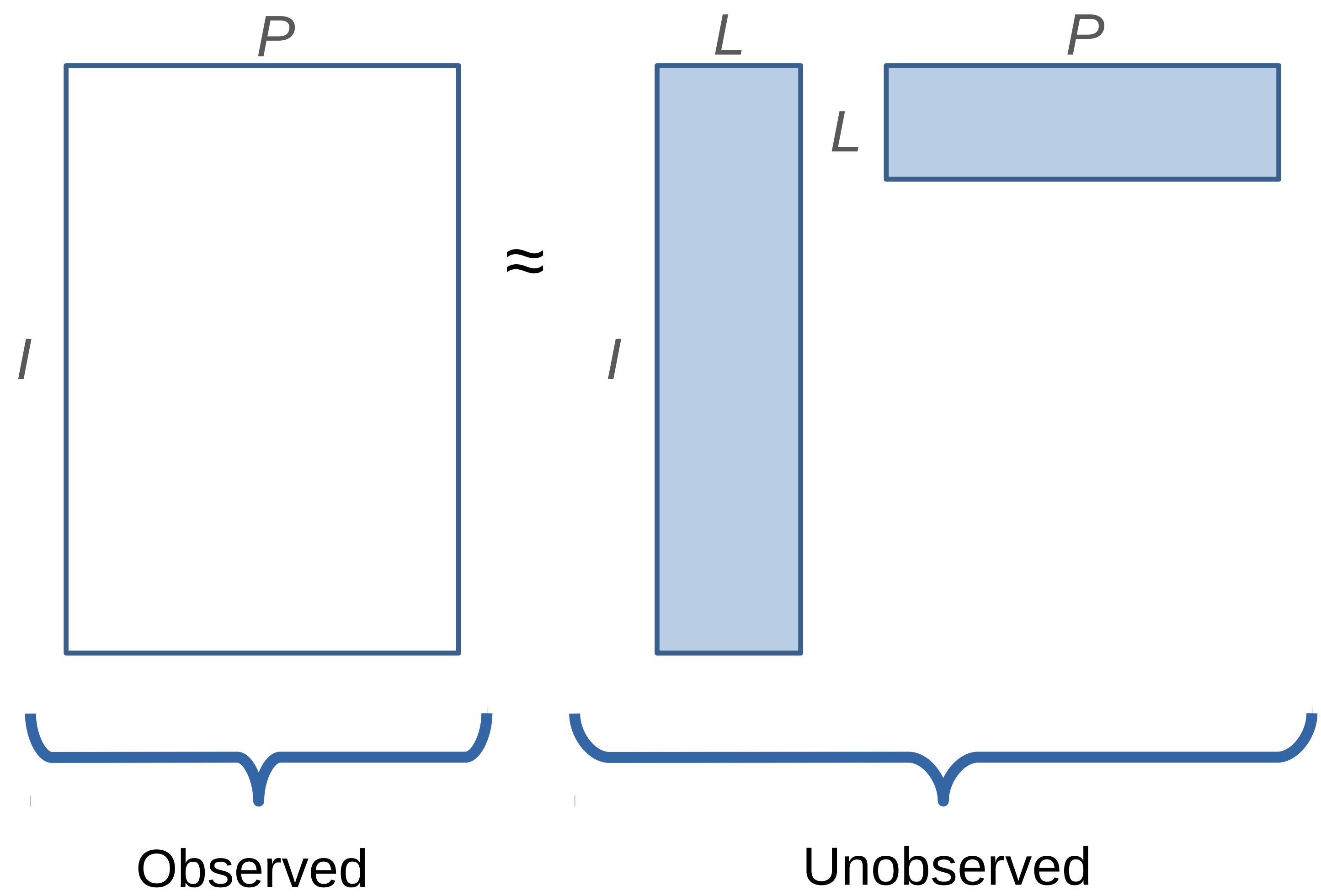


Image: nps.gov

Source apportionment models

$$X \approx F \times \Lambda$$

pollutant concentrations | source concentrations | source profiles



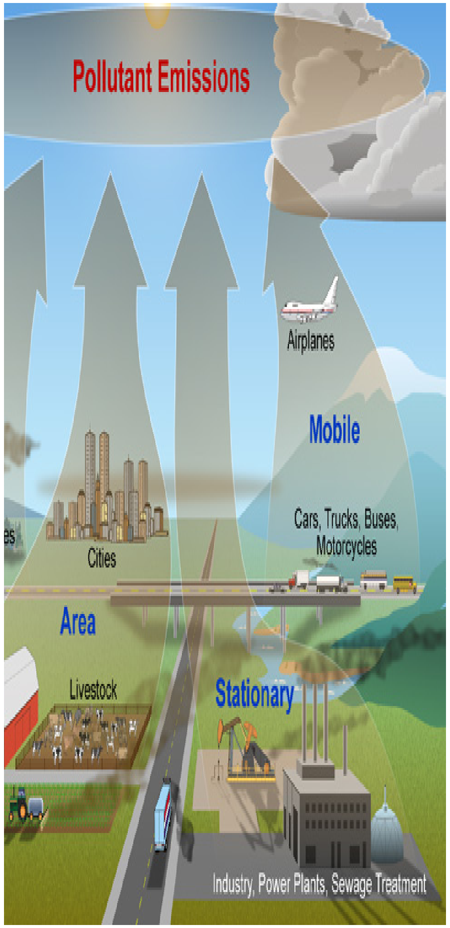
Challenges in source apportionment (SA)

Need to estimate:

- F source concentrations
- Λ source profiles

Challenges

- Choosing number of sources L
- **Deciding on prior information**
  - SA models are not identifiable
  - $X = F\Lambda = F\,RR^T\,\Lambda = F^*\Lambda^*$
- **Naming sources**
  - Understanding major contributors
  - **Attributing health effects**



Positive matrix factorization (PMF)

Find *F* and *Λ* that minimize

$$Q = \sum_{i=1}^{I} \sum_{p=1}^{P} \frac{\left(x_{ip} - \sum_{l=1}^{L} f_{il}\lambda_{lp}\right)^2}{u_{ip}^2}$$

such that $f_{il} > 0$ *and* $\lambda_{lp} > 0$ *for all i,l,p* where

x_{ip} Amount of pollutant *p* for observation *i*

f_{il} Amount of source *l* for observation *i*

λ_{lp} Amount of pollutant *p* in source *l*

u_{ip} Known observation-specific uncertainties corresponding to x_{ip}

- No prior information is needed

Positive matrix factorization (PMF)

- PMF has a graphical user interface
  - Generally finds solutions that minimize Q without additional constraints
  - Run PMF multiple times, find smallest Q
  - Uses the multilinear engine (ME) to solve

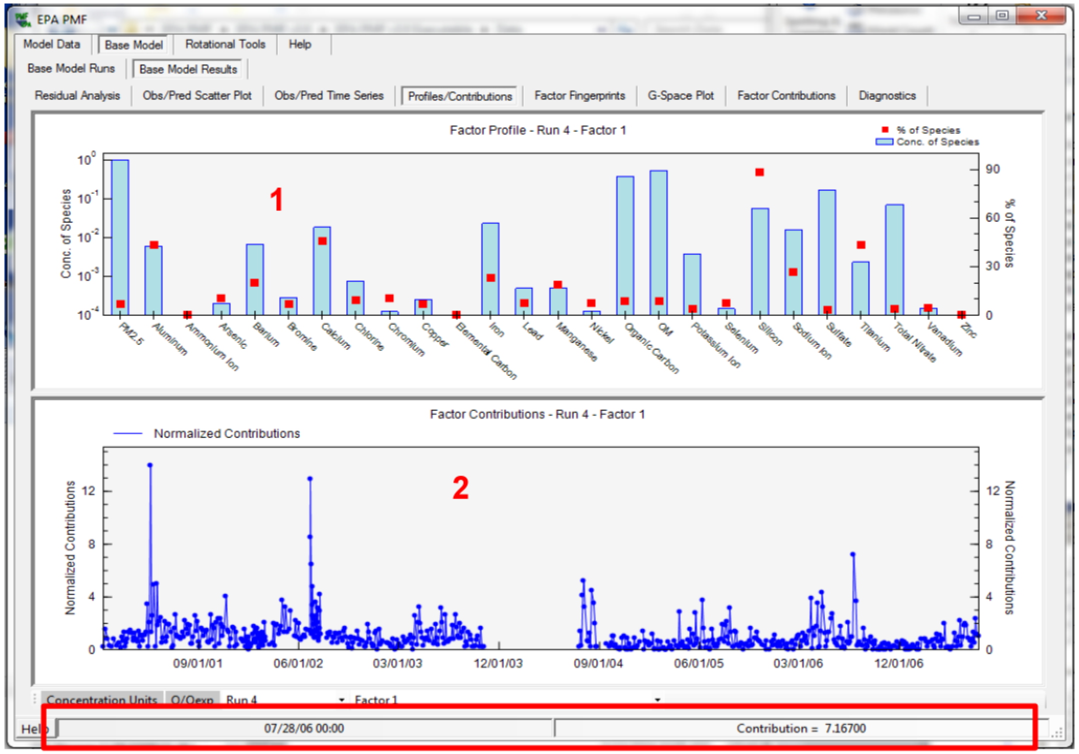


Image: epa.gov

Multilinear engine (ME-2)

ME-2

- more flexible: can incorporate constraints
- minimizes $Q + Q^a$, where Q^a consists of auxiliary equations:

$$Q^a = \sum_{j=1}^{J} Q_j^a = \sum_{j=1}^{J} (r_j / s_j)^2$$

where

r_j is the residual of equation j

s_j is the “softness” of the equation

r_j may be something like $(f_{il} - f^*)$, where f^* is some fixed value.

Health model

Treat estimated source concentrations f_{il} as known in standard health effects regression model:

$$Y_{it} = \beta_0 + \beta_1 f_{il} + \beta_2 z_i + \beta_3 f_{il} z_i + \gamma t + b_i + \epsilon_{it}$$

Y_{it} Difference from baseline for an outcome for observation *i* at time *t*

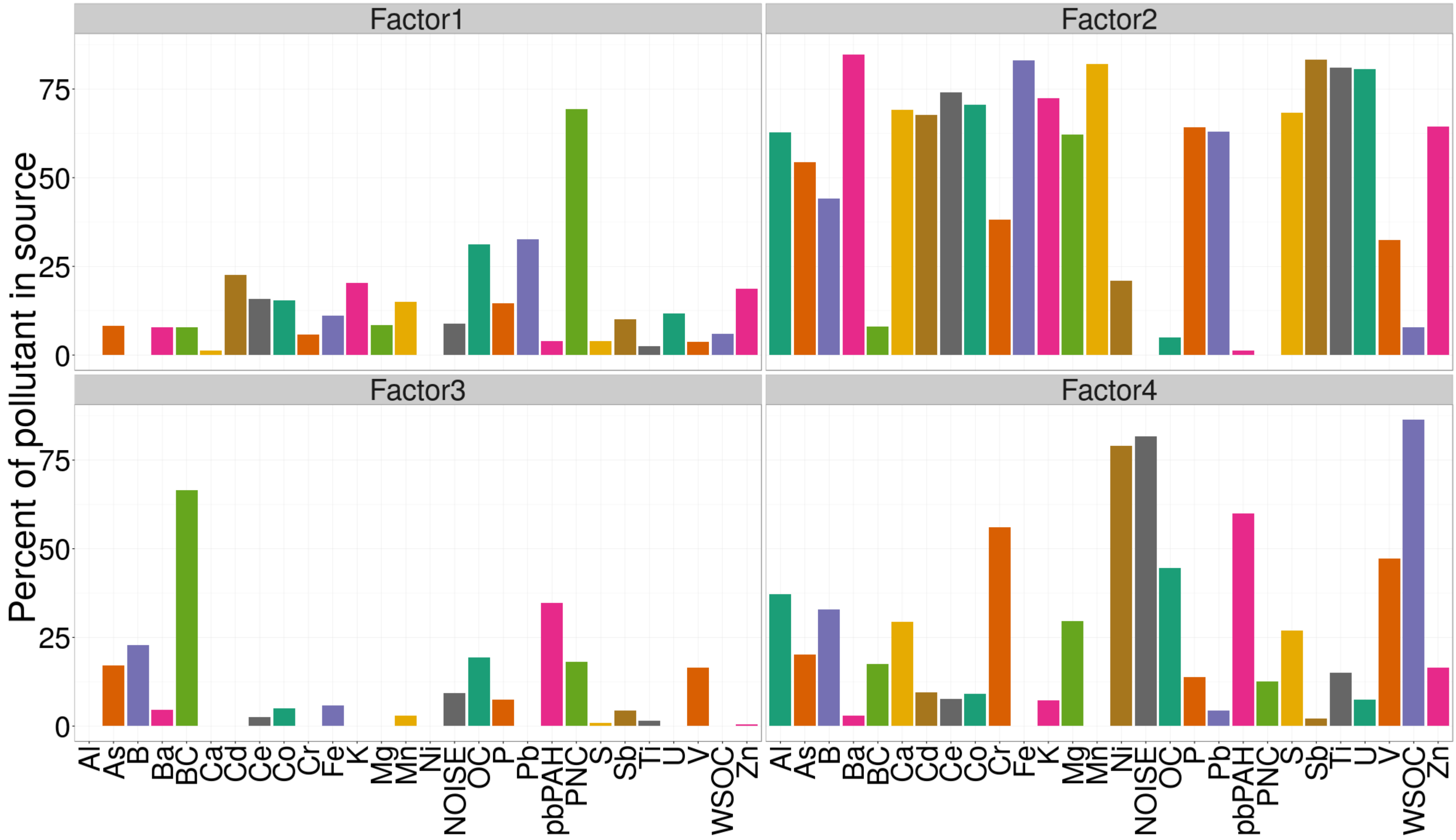
f_{il} *(estimated)* Amount of source *l* for observation *i*

z_i Asthma status (1 = yes) for observation *i*

b_i Random intercept

ε_{it} Random error

Results: source profiles Λ from PMF

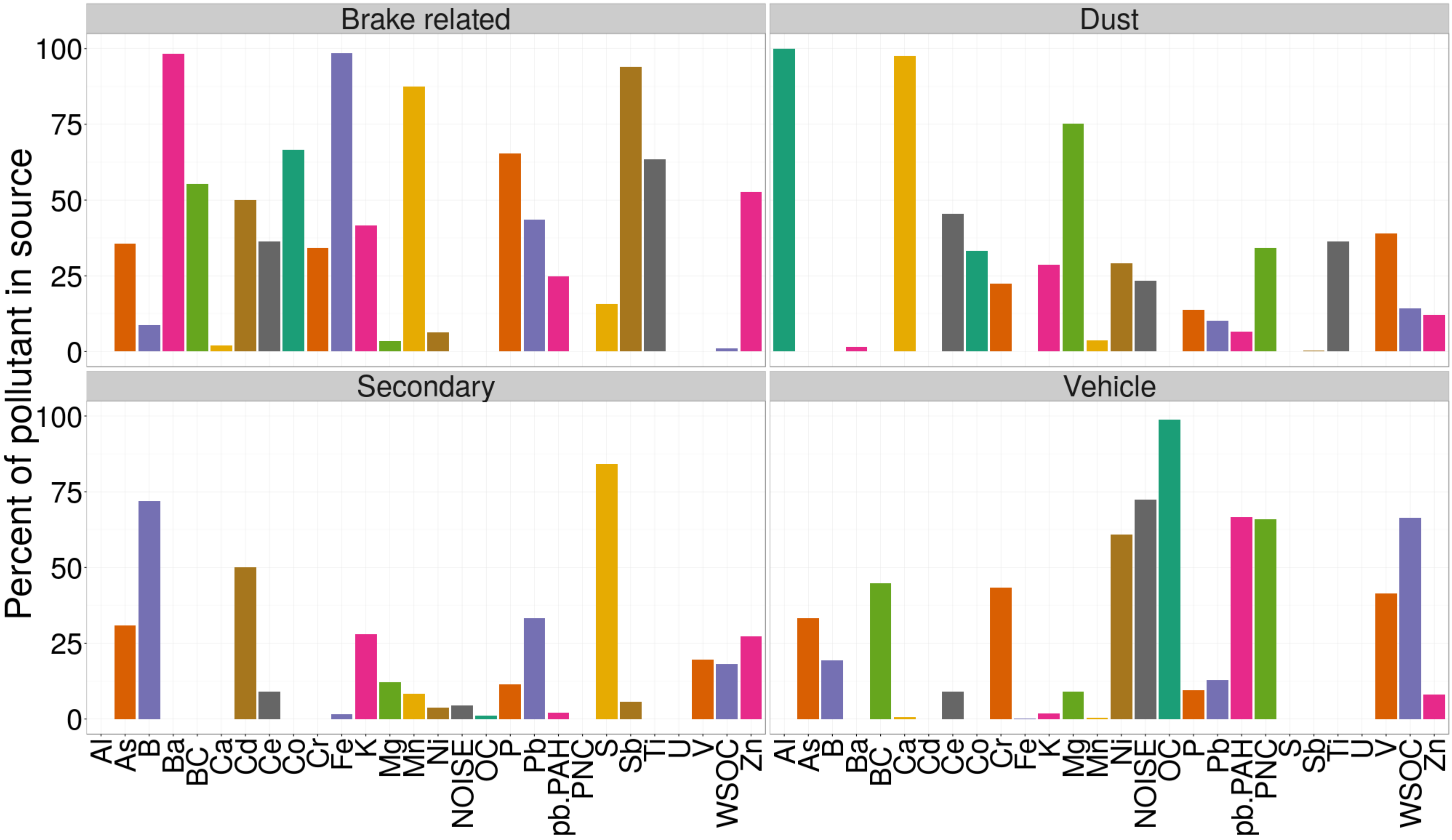


Multilinear engine (ME-2)

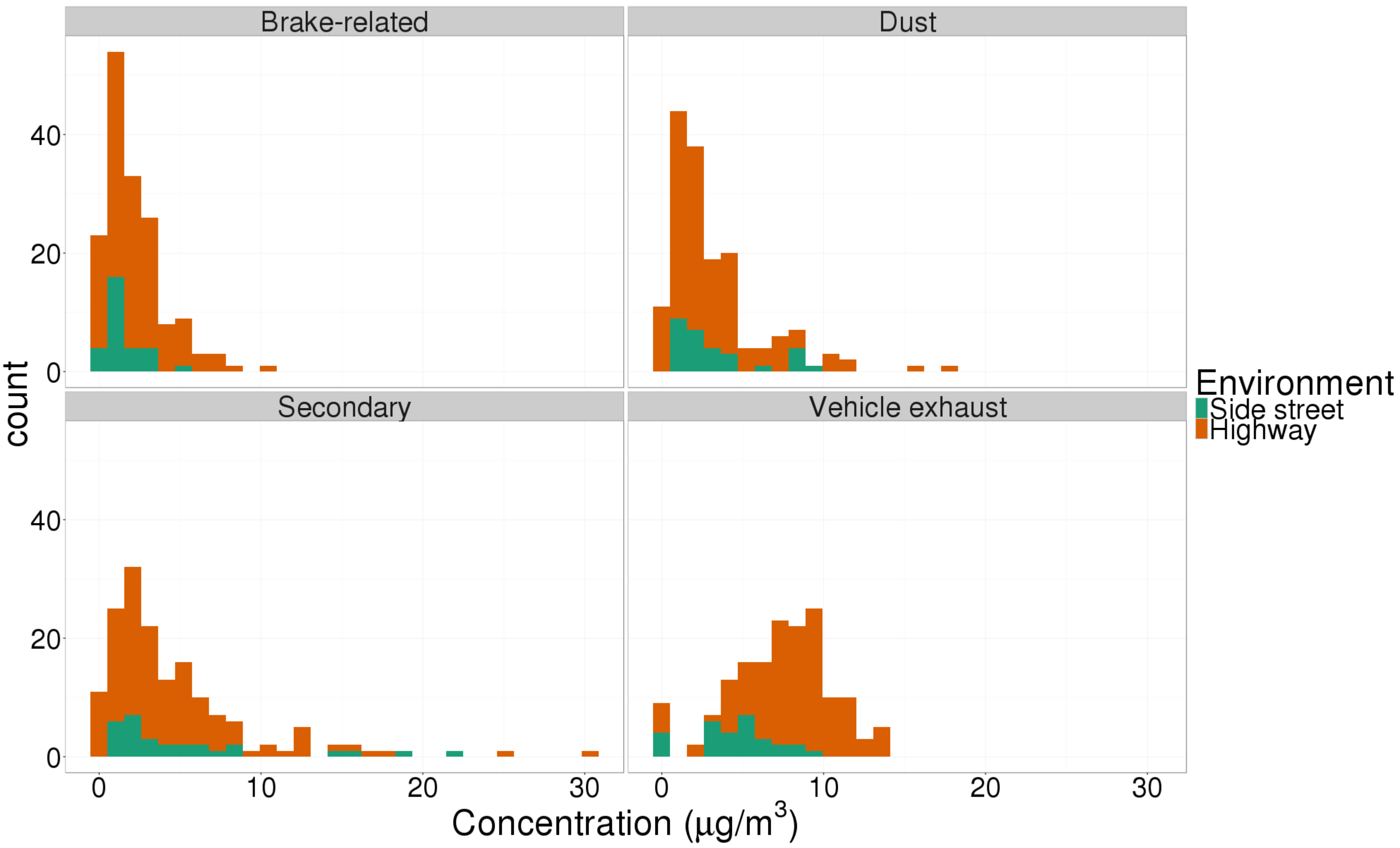
- What auxiliary equations to include?
- Sources:
  - Brake-related (high in metals elements: Zn, Mg)
  - Dust (high in crustal constituents: Al, Ca)
  - Secondary (high in sulfate)
  - Vehicle exhaust (high in carbon-containing constituents: OC, BC)

Source	Brake-related	Dust	Secondary	Vehicle exhaust
Contributing pollutants	Zn, Fe	Al, Ca	S	OC, BC
ME-2 constraints	↑ Zn, Fe	↑ Al, Ca, noise	↑ OC, WSOC	
		↓ WSOC	↓ Al, Ca	↓ OC, BC

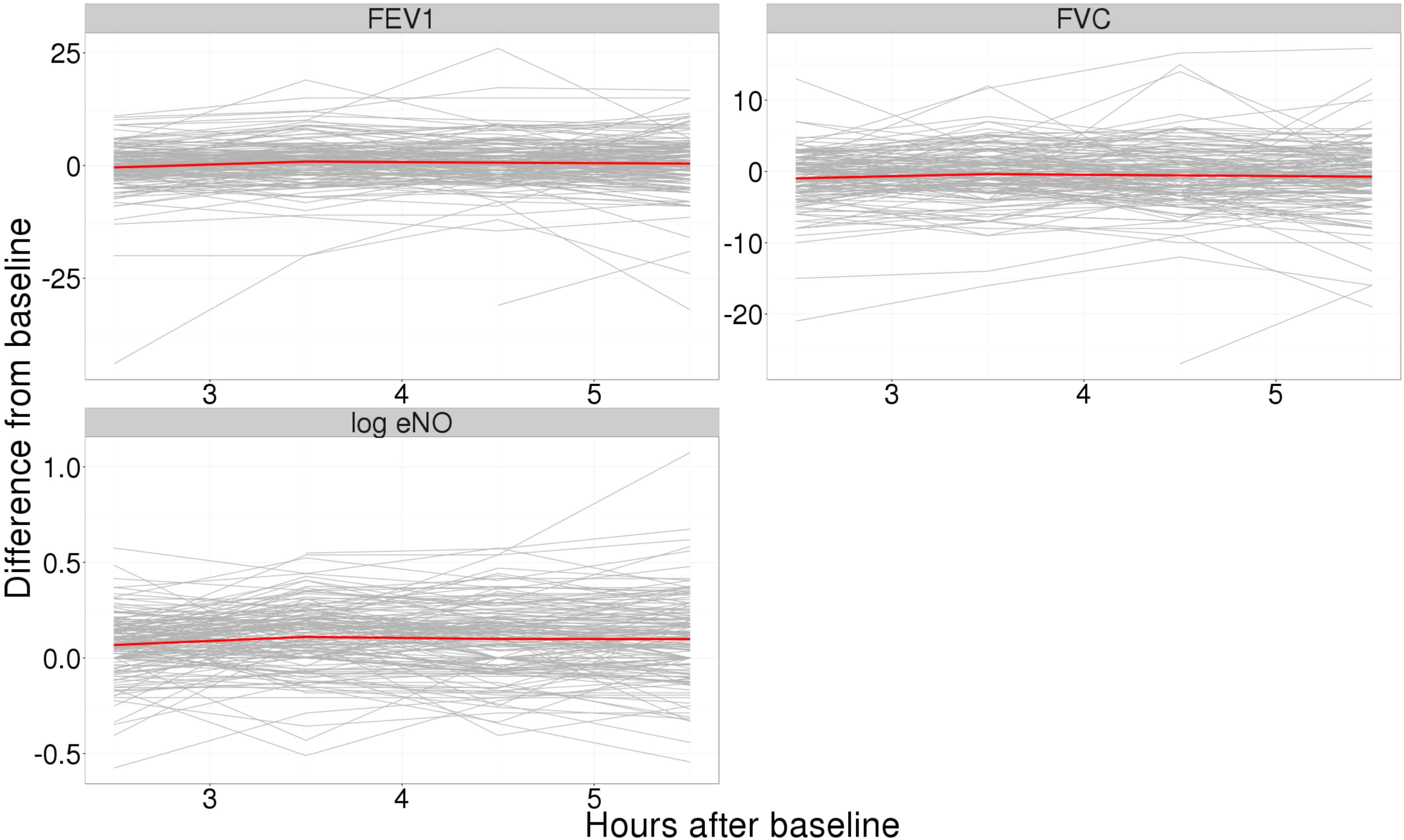
Results: source profiles Λ from ME-2



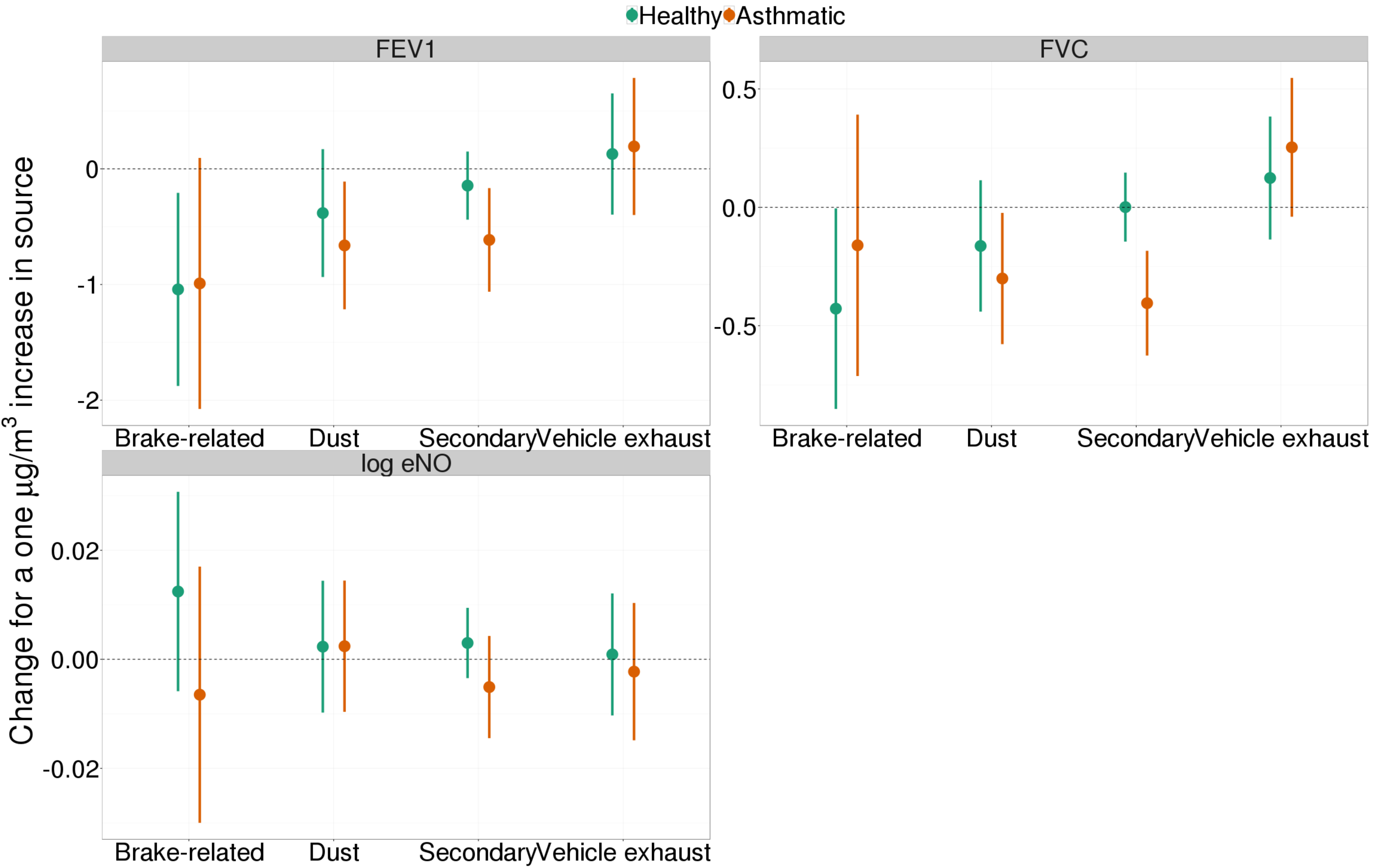
Results: source concentrations F from ME-2



Results: longitudinal health data



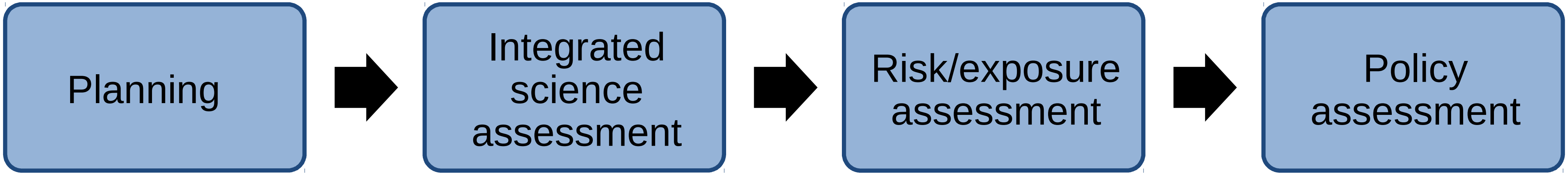
Results: estimated effects of sources on health



Air pollution policies

Some evidence of associations between exposure to secondary pollution and pollution from dust and lung function in individuals with asthma

National ambient air quality standards



Emissions-specific policies

Light-Duty Vehicles, Light-Duty Trucks, and Medium-Duty Passenger Vehicles: Tier 2 Exhaust Emission Standards and Implementation Schedule

Standard	Emission Limits at 50,000 miles					Emission Limits at Full Useful Life (120,000 miles) [2]				
	NOx (g/mi)	NMOG (g/mi)	CO (g/mi)	PM (g/mi)	HCHO (g/mi)	NOx (g/mi)	NMOG (g/mi)	CO (g/mi)	PM (g/mi)	HCHO (g/mi)
Bin 1	-	-	-	-	-	0	0	0	0	0
Bin 2	-	-	-	-	-	0.02	0.01	2.1	0.01	0.004
Bin 3	-	-	-	-	-	0.03	0.055	2.1	0.01	0.011
Bin 4	-	-	-	-	-	0.04	0.07	2.1	0.01	0.011
Bin 5	0.05	0.075	3.4	-	0.015	0.07	0.09	4.2	0.01	0.018
Bin 6	0.08	0.075	3.4	-	0.015	0.1	0.09	4.2	0.01	0.018

Image: epa.gov

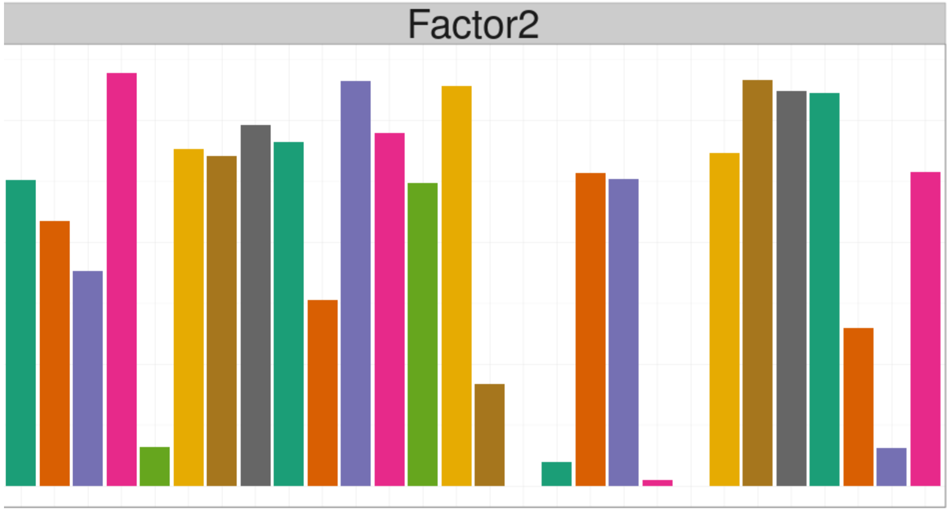
Conclusions

Some limitations:

- Healthy, young population
- Short-term exposure
- Lung function measurements
  - Biomarkers of inflammation
  - Cardiovascular health measures
  -

Evaluating source apportionment models

- Estimated sources may not align well with known sources
- Incorporating prior information
- Validation of results



Thanks!

This research is funded by
U.S. EPA - Science To Achieve Results (STAR) Program
Grant # R834799

jenna.krall@gmail.com

Acknowledgements:

Emory University
Jeremy Sarnat
Lance Waller
Chandresh Ladva

Nankai University
Guoliang Shi
Xing Peng

Georgia Tech
Ted Russell

Ben-Gurion University of the Negev
Rachel Dassa Golan

George Mason University
Department of Global and Community Health

This presentation was made possible by grants from NIEHS (T32-ES012160) and US EPA (R834799). This publication's contents are solely the responsibility of the grantee and do not necessarily represent the official views of the US EPA. Further, US EPA does not endorse the purchase of any commercial products or services mentioned in the publication.

Covariate summary

var	commute	n	% missing	mean	median	sd	min	max
age	0	73	5.2	30.72	28.13	9.03	0.33	59.47
age	1	79	6	30.43	28.33	8.86	0.08	58.44
bmi	0	56	27.3	23.8	23.12	4.08	17.82	38.98
bmi	1	79	6	23.76	23.29	4.08	16.7	38.98
cort_1	0	54	29.9	680.82	453.21	599.37	14.73	2886.51
cort_1	1	61	27.4	742.23	466	622.74	14.73	2886.51